Science

The Macaque Genome

AAAS

CONTENTS

COVER

A rhesus macaque in the NIH Animal Center, Poolesville, MD. The newly available genome sequence of the macaque will enhance its importance as a model for biomedical research and for studies on primate evolution. A special section beginning on page 215 describes the sequencing efforts and some initial fruits of the sequence analysis, and includes a poster, online videos, interviews, and educational tools.

Photo: Joshua Moglia

>> *Editorial p. 173; for related online content, see page 167 or go to www.sciencemag.org/sciext/macaque/*

DEPARTMENTS

EDITORIAL

SPECIAL SECTION

The Rhesus Macaque Genome

INTRODUCTION

NEWS

RESEARCH ARTICLE

REPORTS

Special Pullout Feature >> Poster: The Macaque Genome

188

NEWS OF THE WEEK

NEWS FOCUS

Science

211, 272, &
274

REPORTS

206 &
285

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

SCIENCE**NOW**

www.sciencenow.org DAILY NEWS COVERAGE

Reflections of Absolute Zero
Researchers freeze a small mirror to within one degree
of witnessing quantum effects.

Forest Elephants on a Road to Ruin
Poachers use logging roads to pursue the endangered
animals.

Look Out for Alien Lasers
Astronomers are trying to see light from extraterrestrials
using gamma-ray telescopes.

Impact on American postdoc pay.

SCIENCE **CAREERS**

www.sciencecareers.org CAREER RESOURCES FOR SCIENTISTS

US: Huddled Masses
B. Benderly
The large influx of international postdocs may not be good
for the incomes of America's young scientists.

GLOBAL: Mastering Your Ph.D.—Science Papers That Shine
P. Gosling and B. Noordam
Part of being a good scientist is being able to write up your results
in clear and simple terms.

US: Opportunities—A Day in the Life, Part 2
P. Fiske
It pays to work hard to find people who can help you.

US: Negotiating—Please Sir, Can I Have Some More?
V. Mohan-Ram
An article from the archives offers advice on negotiating a job offer.

Neurite outgrowth of PC12 cells.

SCIENCE'S **STKE**

www.stke.org SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

**PERSPECTIVE: Regulation of PC12 Cell Differentiation
by cAMP Signaling to ERK Independent of PKA—Do All
the Connections Add Up?**
M. J. Gerdin and L. E. Eiden
The exact protein kinase A–independent path that is involved
in neurite formation in PC12 cells remains elusive.

**FORUM: Response to "Can Mesoscopic Models Test
Spatial Mechanisms of Cell Signaling?"**
U. Bhalla
Simulation methods and available data encourage optimism
about modeling spatiotemporal signaling in and under the
membrane.

SCIENCE**ONLINE FEATURE**

POSTER: The Macaque Genome
An interactive version of the pullout poster
in this issue.
www.sciencemag.org/sciext/macaqueposter/

Separate individual or institutional subscriptions to these products may be required for full-text access.

<< A Double-Tethered Switchblade

Fatty acids, which are comprised mainly of long hydrocarbon chains and serve essential structural and energetic functions in cells, are synthesized by adding two-carbon building blocks to a starter unit. Each of the additions involves a series of four reactions; for example, synthesis of a palmitate chain requires cycling seven times through this set of four catalytic sites. **Jenni** *et al.* (p. 254) and **Leibundgut** *et al.* (p. 288) describe the crystal structures of the fatty acid synthase complexes from the fungus *Thermomyces lanuginosus* and the yeast *Saccharomyces cerevisiae*. For the fungal enzyme, a complete mapping of the catalytic domains within the two-chambered heterododecameric ($a_6 b_6$) complex is provided. The yeast data reveals the cyclical path taken by the acyl carrier protein (ACP) domain to which the nascent fatty acid is attached. The ACP moiety is tethered to the wall and to the floor of the chamber, which constrains its movements as it visits the nearby four catalytic sites. Upon arrival, it unfolds the growing acyl chain like a switchblade.

Proton Tug-of-War

In acidic aqueous solutions, protons are shared and shuttled by the solvent molecules or dissolved bases, as opposed to moving about as free H^+ ions. Probing such structures is challenging, however, because the many energetic configurations that form at ambient temperature lead to very broad spectral bands. **Roscioli** *et al.* (p. 249) have used gas-phase argon clusters to isolate and probe the vibrations of complexes in which a proton bridges two molecules of widely varying basicity, ranging from water and ammonia to alcohols, ethers, and noble gases. The infrared spectra of these cold complexes show sharp absorption bands that clarify how the proton affinities and skeletal vibrations of the flanking bases impact the motion of the H^+ ion confined between them.

All-Organic Frameworks in Three Dimensions

Numerous metal-organic framework compounds have been reported in which high surface areas are achieved by the metal centers directing the assembly of linking organic groups. **El-Kaderi** *et al.* (p. 268; see the Perspective by **Budd**) now report the synthesis and structural characterization of high-surface-area, covalent organic frameworks through the condensation of subunits that can form four bonds tetrahedrally with another type of subunit that can form three bonds triangularly. After target networks were chosen, molecular design programs were used to optimize the choice of subunits. The strong covalent bonds in the framework (C–C, C–O, C–B, and B–O) lead to high thermal stability (400° to 500°C), and the use of only light atoms leads to low densities (0.17 grams per cubic centimeter).

Nanoparticles Take Shape

Ceramics are often made from "greenwares," in which aggregates of small colloidal particles are molded or shaped before thermal reactions
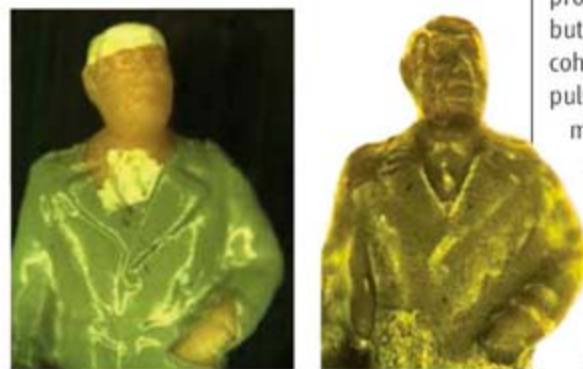
remove solvent and bond the particles together. **Klajn** *et al.* (p. 261) show that metal nanoparticles (NPs) can be similarly molded into macroscopic objects. The metal NPs are coated with a surfactant that can undergo ultraviolet-induced isomerization from a trans to cis configuration. The higher dipole of the cis form causes the NPs to aggregate into larger "superspheres" 50 to 300 nanometers in diameter. These superspheres adhere to each other and allow formation of shapes and coating of objects (such as small figurines). Subsequent annealing creates hardened polycrystalline porous materials that can be made from either single or mixed metal NPs.

Raman Probes Shape Up

Raman spectroscopy can provide a wealth of information about molecular vibrations and provide fingerprint signatures for identification, but even when signal strength is boosted by coherently exciting the vibrations with laser pulses, a fluctuating background signal hinders many practical applications in sensing. **Pestov** *et al.* (p. 265; see the Perspective by **Lucht**) now describe a method in which a probe pulse is delayed and has its shape optimized so to minimize the nonresonant background contributions. The authors apply this method to the detection of picolinates, the characteristic component of bacterial spores such as anthrax.

A Light Touch for Spin

Differences in the pressure of warm sunlight being reflected and re-radiated from the surface

of an asteroid during its orbit can change how it spins. This process, called the Yarkovsky-O'Keefe-Radzievskii-Paddack (YORP) effect, has been predicted but not seen directly. Two reports describe the detection of the YORP effect acting on the near-Earth asteroid 54509 (2000 PH5); see the Perspective by **Rubincam and Paddack**. **Lowry** *et al.* (p. 272, published online 8 March) monitored the reflected optical light from the asteroid to show how the spin rate of the asteroid is decreasing. **Taylor** *et al.* (p. 274, published online 8 March) have mapped the asteroid's shape using radar observations to show that this slowing is precisely as predicted by the YORP effect.

Ancient Collagen Signatures

Soft tissues have been thought to be rarely if ever preserved in the fossil record, aside from some samples entombed in amber or for a few million years in ice. Recently, a femur of a *Tyrannosaurus rex* dating to about 67 million years ago was recovered that seemed to preserve internal soft tissues, including blood vessels within its bone. **Schweitzer** *et al.* (p. 277) and **Asara** *et al.* (p. 280) have further analyzed these tissues, as well as samples from a mastodon, and show that original collagen proteins were preserved. Mass spectrometry was used to recover at least some of the original collagen sequence. Thus, aspects of genetic information can be obtained from select samples of extinct species preserved for tens of millions of years.

Spotlight on the Pre–B Cell Receptor

The pre–B cell receptor (pre-BCR), comprising a heavy chain and a heterodimeric surrogate light chain (SLC), a signaling complex that acts as a checkpoint in B cell development. **Bankovich** *et al.* (p. 291) report the structure of a pre-BCR Fab-like fragment at 2.7 angstrom resolution. The structure shows how the requirement for pairing with the SLC might constrain the repertoire of heavy chains in the mature antibody population. The crystal structure, together with electron microscopy data and biochemical analysis, supports a model of antigen-independent, SLC-mediated dimerization of the pre-BCR to promote pre–B cell activation and expansion.
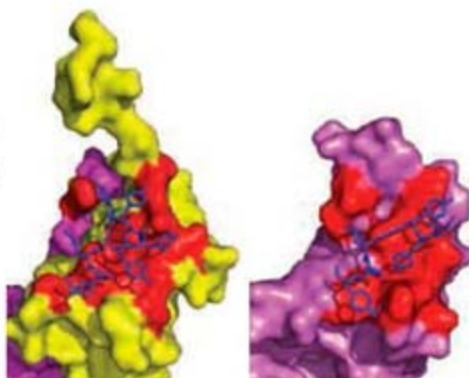
Making LIGHT of Lipid Metabolism

Atherosclerosis results from a combination of lipid dysregulation and inflammation-mediated pathology of the vasculature. **Lo** *et al.* (p. 285; see the Perspective by **Hansson**) show that increased expression of related members of the tumor necrosis factor family of inflammatory cytokines, LIGHT and lymphotoxin (LT), on T cells can elevate circulating blood cholesterol and triglycerides in mice. This effect appeared to be mediated via lymphotoxin β receptor (LTβR) signaling in hepatocytes, leading to a drop in the activity of hepatic lipase, an enzyme central to lipid metabolism. The normally high lipid levels found in mice that lack the low-density lipoprotein receptor gene were reduced when LTβR signaling was inhibited. These results raise questions about how the immune system detects and subsequently exacerbates dyslipidemia, and whether this process makes any direct contribution to atherosclerosis in humans.

Double Source for S1P

Sphingosine-1-phosphate (S1P) is a circulating lipid mediator that induces the egress of lymphocytes from lymphoid organs. The immunomodulatory effects of S1P are made apparent by the absence of circulating lymphocytes in mice that are unable to support its production and by the encouraging results of clinical trials aimed at targeting this pathway to suppress transplant rejection and autoimmunity. **Pappu** *et al.* (p. 295, published online 15 March; see the Perspective by **Chun**) use a combination of conditional gene deletion and bone marrow chimerism to illuminate two sources of S1P in the blood and lymphatic circulation. By sustaining S1P levels outside the lymphoid organs, these supplies allow lymphocytes to follow a gradient between the lymphoid tissue—where S1P is catabolized to low levels—and the two circulatory systems. This insight may help refine approaches of immune suppression and activation via the S1P pathway.

CREDIT: BANKOVICH ET AL.

Moving the Primate Debate Forward

IN THE UNITED KINGDOM, PUBLIC CONCERN ABOUT THE USE OF ANIMALS FOR research has a long and checkered history. In 1875, Charles Dodgson, better known by his pseudonym Lewis Carroll as the author of *Alice's Adventures in Wonderland*, wrote a fierce polemic on vivisection in an attempt to prevent the establishment of a physiology department at Oxford University. The activities of animal rights movements have since reached new dimensions, ranging from threatening mail and personal violence to letter bombs and worse. Nevertheless, opinion polls show that the majority of the UK public accepts the need to use animals for medical research. What they are less happy about is the use of primates, particularly for what is perceived as curiosity-driven research rather than work with a medical objective. The debate on this topic is likely to remain highly controversial in the United Kingdom, but a recent report* by an independent group of scientists and nonscientists outside the primate research community attempts to provide a better-informed basis for this debate through an in-depth analysis of the scientific reasons for research on monkeys. Most important, it calls for a national strategic plan for nonhuman primate research. The sponsors of the report—the Royal Society, Medical Research Council, Wellcome Trust, and Academy of Medical Sciences—are expected to respond to the report's recommendations by June 2007.

Because no great apes have been used for research in the United Kingdom since 1986, the report deals mainly with the use of monkeys in basic or applied research, making the case that modern biomedical research encompasses a continuum between them. It focuses on the neurosciences and on communicable diseases, particularly the development of vaccines for HIV/AIDS, tuberculosis, and malaria. For each case, global health importance and approaches that could be taken to avoid the use of animals are examined in detail. Other research areas are considered briefly, as well as ethics, animal welfare, drug discovery, and toxicology. The current status of alternatives to animal use is also reviewed, ranging from molecular, cellular, and noninvasive approaches for studying the nervous system, to stem cell biology and pharmaco-metabonomic phenotyping.

The report concludes that in some cases there is a valid scientific argument for the continued use of monkeys. Although the amount of biomedical research done in the United Kingdom has almost doubled over the past 10 years, the number of monkeys used has remained relatively constant, indicating that alternative research venues are being pursued. However, because of the speed of development in the biomedical sciences and the increasing availability of alternatives to animal use, no blanket decisions can be made. Rather, each case must be considered individually, supported by a fully informed assessment of the importance of the work and of approaches that do not require animals. To this end, the national strategic plan for primate research called for by the report includes the regular dissemination of information about alternative methods and the creation of centers of excellence, both for the better care of animals and for the training of scientists. The plan also emphasizes openness by journals in disclosing the suffering caused to animals and calls for regular publication of the outcomes of primate research and toxicology studies by funding bodies and the pharmaceutical industry.

Over recent years, the UK government has taken steps to protect scientists and others who are involved in animal research. We hope that it will now join forces with the sponsors of this report to activate its recommendations. The public debate on nonhuman primate research needs to move forward on the basis of sound scientific reasons. The increasing study of biology and disease at the cellular and molecular levels, supported by small-animal models, will probably reduce the requirement for primates in research. However, we do not currently know the most effective approach in some vital areas. Thus, it would be extremely unwise to rule out primate use for the foreseeable future.

– David Weatherall and Helen Munn

*The Use of Non-Human Primates in Research (www.acmedsci.ac.uk/images/project/nhpdownl.pdf).

David Weatherall chaired the group that produced this report and is Regius Professor of Medicine Emeritus at the University of Oxford, UK.

Helen Munn is Director of Medical Science Policy at the Academy of Medical Sciences, London, UK.

Gypsum megacrystals.

GEOLOGY

Very Slow Growth

Gypsum [Ca(SO$_4$)·2H$_2$O] forms some of the largest natural single crystals on Earth (aside from the speculative iron crystals in the inner core), in some cases reaching 10 m in length. The growth of such sizable crystals requires precise maintenance of specific environmental conditions.

Garcia-Ruiz et al. have investigated the giant gypsum crystals in deep caves of the Naica mine in Mexico, which has been the source of several museum specimens. Analyses of fluid inclusions, trapped sequentially in the crystals as they grew in caverns nearly 300 m below the surface, show that the temperature in the large fluid-filled caves was maintained near 54°C for thousands of years at least—the mineralization in the mine began about 25 million years ago—and the deep water there is still close to this temperature today. This temperature is just below the maximal solubility point for gypsum in low-salinity water and also slightly below the thermodynamic stability range of anhydrite (a polymorph of gypsum), which had formed previously. Thus, the dissolution of anhydrite maintained a slight supersaturation of gypsum in the fluid, and a temperature close to the equilibrium allowed the formation of only a few crystal nuclei in the deep large cavities. Shallower, cooler cavities have produced multiple smaller crystals. — BH

Geology **35**, 327 (2007).

APPLIED PHYSICS

A Peek Inside

The semiconductor industry routinely fabricates device structures with feature sizes smaller than 100 nm. With millions of components crowded onto each chip and complex circuitry arrayed in three dimensions, methods to test the structures for defects— preferably nondestructively and with high throughput—become challenging. Techniques for imaging the subsurface structures tend to face a tradeoff between resolution and contrast. The probe light must have a relatively long wavelength (usually in the infrared) in order to penetrate through several millimeters of silicon in the wafer and be absorbed by the active layers of the device; however, this wavelength requirement has generally restricted lateral resolution. Ramsay et al. combine immersion lens imaging with two-photon absorption microscopy to improve both the lateral resolution and the absorption contrast, thereby providing a technique for the high-resolution three-dimensional imaging of the subsurface structures in silicon chip circuitry. — ISO

Appl. Phys. Lett. **90**, 131101 (2007).

BIOCHEMISTRY

A Nanomembrane

The technical difficulties of working with membrane proteins, which sport extensive hydrophobic and hydrophilic surfaces (not to mention a hetero-geneous collection of attached sugars), are matched only by the ease with which cells manage to handle them. In bacteria, the trimeric complex SecYEG accepts substrate proteins made in the cytoplasm and either passes them through the inner membrane to the periplasmic space or ejects them laterally straight into the inner mem-
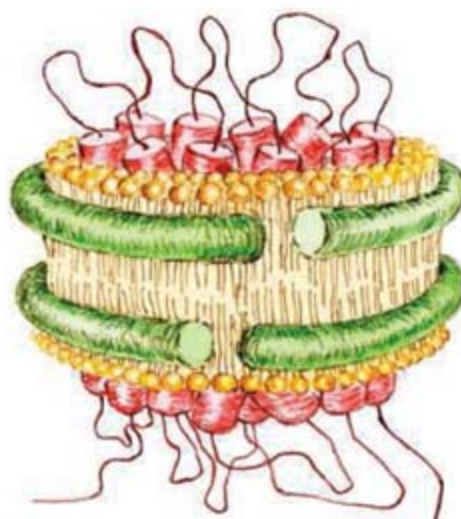
SecY (red) in a lipid (yellow)/protein (red) matrix.

brane itself (how outer membrane proteins are dealt with is a whole other story). Some of the substrates are delivered by the cytosolic motor protein SecA, but the amphiphilic character of the protein translocation machinery has made it hard to probe the structural state of functional SecA-SecYEG interactions. Alami et al. have reconsti-tuted SecYEG monomers into a membrane-like lipid/protein construct, referred to as a nanodisc; adding dimeric SecA to these nanodiscs results in dissociation of the dimers and binding of monomeric SecA to SecYEG and the consequent stimulation of SecA ATPase activity. — GJC

EMBO J. **26**, 10.1038/sj.emboj.7601661 (2007).

MATERIALS SCIENCE

Approaching the Ideal

Frenkel predicted 80 years ago that the ideal strength of a metal should be 1/5 of its shear modulus, but in most metals the actual strength ratio is closer to 1/1000 because of the motion of dislocations at much lower stresses. Li et al. use computational methods in an effort to understand the behavior of a family of body-centered cubic (bcc) Ti-Nb–based alloys known as Gum Metals. These alloys have the unusual property of sustaining very large elastic deformations before yielding, as well as substantial plastic deformation before failing. The authors argue that for this behavior to occur, the ideal strength must be below a stress at which the material would deform by ordinary dislocations, and that the material must always fail by shear rather than cleavage fracture. Using ab initio calculations to determine the elastic properties of related Ti-V alloys, they find that at a ratio of valence electrons to atoms close to the Gum Metal value, the bcc lattice becomes unstable; thus, the Gum Metals intrinsically have a low ideal strength and

tend to fail in shear even when pulled in tension. Further, at values close to this transition, it is possible to introduce sufficient obstacles for dislocation motion through the addition of extra alloy elements without complete loss of ideal strength. The authors believe that similar computations could identify useful alloys that exist close to this edge of bcc stability. — MSL

Phys. Rev. Lett. **98**, 105503 (2007).

BIOMEDICINE

Looking for Cancer Stem Cells

The intense interest in stem cell research has helped to revive the cancer stem cell hypothesis, which postulates that tumor cell growth is driven by a small population of malignant cells that have the ability to self-renew and to differentiate—a capacity that is shared with normal tissue stem cells. The idea is attractive because it suggests that drugs could be designed to target cancer stem cells selectively, if and when these cells are identified. Although the stem cell origin of leukemias is now widely acknowledged, the role of stem cells in solid tumors has been more contentious. Shipitsin *et al.* performed a comprehensive molecular characterization of two classes of cells purified from human breast cancer: one class

A network of genes up-regulated in normal (blue) or cancer (red) CD44+ cells.

expressed a cell surface marker (CD44) previously associated with high tumorigenicity and stem cell–like properties, and the second class expressed a marker (CD24) previously associated with low tumorigenicity and a more differentiated state. The CD44+ breast cancer cells were found to express many genes in common with progenitor cells in normal breast tissue, and the abundance of these cells in the tumor appeared to correlate with decreased patient survival. However, the CD44+ and CD24+ cells within individual breast tumors showed genetic differences, a finding that does not fit neatly with the simplest version of the cancer stem cell hypothesis. An alternative model is that many cancer cells retain the capacity to adapt to changing conditions, whether this means reverting to a more primitive, stemlike state or evolving into a more differentiated state. — PAK

Cancer Cell **11**, 259 (2007).

MICROBIOLOGY

A High-Fiber Diet

In the race to replace fossil fuels with biofuels, microbial fermentation may become a key technology. However, microbes can do only so much and balk when their food contains too much lignin. This is not uncommon because the fibrous tangle of lignin and cellulose, called lignocellulose but better known as wood, is ubiquitous. To add to the problem, the enzymatic breakdown of cellulose is not as rapid as the enzymatic breakdown of starches. Jeffries *et al.* present the genome sequence of the yeast *Pichia stipitis* Pignal, which can digest lignocellulose and can transform xylose, a component of lignocellulose, into ethanol. The yeast sequenced was isolated from insect larvae and is related to yeasts found in the gut of beetles that frequent rotting wood. The 15.4-Mb genome is divided into eight chromosomes and includes 5841 predicted genes, including a group of cellulases and xylanases and a number of genes encoding putative xylose transporters. Further analysis showed which genes in which metabolic pathways respond to changes in xylose, glucose, or oxygen. Unlike *Saccharomyces cerevisiae*, which regulates fermentation according to glucose availability, *P. stipitis* regulates fermentation according to oxygen levels, which is reflected in how the genes respond to oxygen. — PJH

Nat. Biotechnol. **25**, 319 (2007).

GENETICS

Networking with Your Peers

Phenotypes embody genotypes, but identifying the steps from coding region to phenotypic variant is not always straightforward because it can often involve complex or multiple protein interactions, or both. These interactions can be decomposed into the direct regulation of genes through protein-protein, protein-DNA, and DNA modifications such as methylation and an indirect regulation that includes genetic interactions between regulator genes. By creating strains of yeast carrying single or double mutations in five transcription factors known to affect filamentous growth and examining their phenotypes and gene expression profiles, Carter *et al.* employed a systematic strategy for generating a model that could be used to estimate phenotypic variation resulting from the mutation of a gene within a network. As a result of accounting for both direct and indirect genetic effects, the authors were able to predict the expression levels of the double mutants on the basis of the single mutants, and to infer functional cross-influences between previously unidentified interactions. — LMZ

Mol. Syst. Biol. **3**, 10.1038/msb4100137 (2007).

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

Remarkably simple system.
Simply remarkable results.

1. Set-up

3. Analyze

2. Run

NEW StepOne™ Real-Time PCR System

The new StepOne™ System makes it easy to get high-quality real-time PCR results.

The StepOne System combines plug-and-play convenience with the uncompromising performance that has made Applied Biosystems the leader in real-time PCR. Whether you're a first-time user or a real-time PCR expert, the StepOne system delivers the highest quality results using the full range of TaqMan® Gene Expression and SNP Genotyping assays. The system's software guides new users through set up, operation and analysis, while also providing power and flexibility to meet the expectations of even the most experienced users. The StepOne Real-Time PCR System is part of an optimized system solution for real-time PCR that includes software, reagents, and instrumentation— all for a remarkably low price.

- Runs intuitive software that teaches you without tutorials or manuals
- Configured for PC-free or networked operation
- Fast runs (<40 minutes) and standard runs (<2 hours) with one block
- Preloaded and precalibrated to get you up and running fast
- Small, space-saving footprint

To learn more about the StepOne Real-Time PCR System or to view a software demonstration, visit info.appliedbiosystems.com/stepone

AB Applied Biosystems

Picturing the Cell

NET WATCH

In an early *Drosophila* embryo, the cell nuclei twirl and divide with the impeccable synchrony of dancers in a Hollywood musical. A lengthwise cut through two sperm tails shows mitochondria lined up like kernels in an ear of corn (below). Those are a couple of the highlights from this gallery hosted by the American Society for Cell Biology in Bethesda, Maryland.

The videos and electron micrographs have all been peer-reviewed to make sure they are scientifically valuable. Included are descriptions of what they illustrate and how they were taken.

The gallery boasts a slew of historic shots from society founders such as the Romanian-American scientist George Palade, now 94, who shared a 1974 Nobel Prize for helping to reveal the internal structure and workings of the cell. Curator David Ennist encourages other biologists to contribute footage and images. >>
cellimages.ascb.org
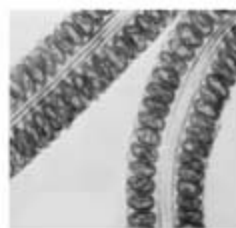
The Whales of Italy

It's been a good week for Italian whales—the ancient, fossilized kind, that is. First, researchers discovered a 4-million-year-old whale skeleton near Pisa. Then amateur paleontologists unearthed the 10-meter-long skeleton of an ancient whale under the vineyards of Castello Banfi, some 55 kilometers from the coast of Tuscany. Analysis of surrounding rocks by Michelangelo Bisconti of the Museum of Natural History of the Mediterranean in Livorno suggests that the latest fossil (below) is about 5 million years old. If carbon-14 dating confirms the age, says paleontologist Lorenzo Rook of the University of Florence, the whale "could cast light on a still-mysterious period" known as the Messinian salinity crisis—6 million years ago—when the Mediterranean Sea largely dried up and then reflooded as water poured back through the Strait of Gibraltar. All of Tuscany was underwater until 2.5 million years ago, when complex geologic forces raised the Apennine mountains and squeezed the region out of the sea.
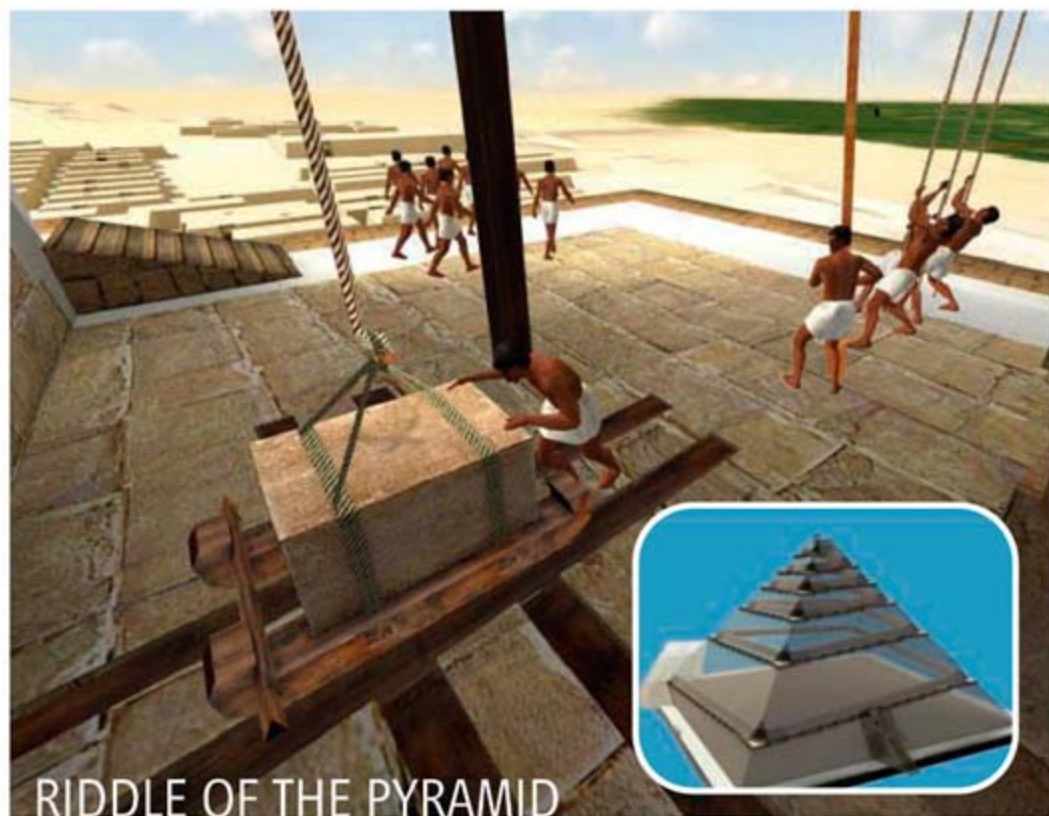
RIDDLE OF THE PYRAMID

A French architect says he has uncovered the secret to the construction of Egypt's Great Pyramid of Cheops 4500 years ago: Workers hauled the stones up an internal spiral ramp.

Jean-Pierre Houdin has been working on his insight for 8 years, and late last month in Paris, he unveiled it along with a video made using new 3D-visualization software.

Houdin says the usual theories of how pyramids were constructed are impractical: A giant ramp would use more stones than the pyramid itself, and a ramp spiraling up the outside would make it hard for engineers to get the geometry right. But a 2-meter-wide inner ramp solves all the problems, he says. Corners of the pyramid would have been left open, allowing workers to maneuver 2-ton blocks around them (see illustration). Houdin is negotiating with Egyptian authorities to allow noninvasive testing of his idea using microgravimetry and infrared and acoustic sensing.

The work was done in consultation with Egyptologist Robert Brier of the C. W. Post Campus of Long Island University in Brookville, New York, who says, "it's a radical new theory, [but] almost all the Egyptology experts say it should be tested." At least one native Egyptian has reservations about it, however: Farouk El-Baz, head of Boston University's Center for Remote Sensing, says, "No engineer would ask workers in ancient Egypt" to haul stones up the dim inner ramps. "These are people that live all their lives in the sun, and most are afraid of the dark."

Racing With The Turtles

Close to 95% of leatherback turtles in the Pacific have disappeared in the past 2 decades. The Costa Rica population has decreased to fewer than 100.
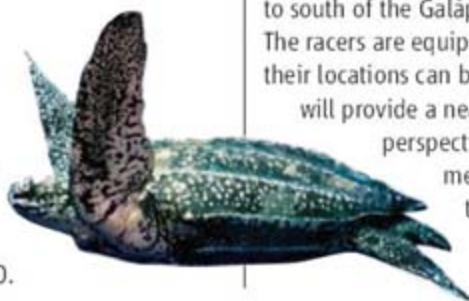
To raise support for the critically endangered beasts, several conservation organizations have created The Great Turtle Race. From 16 through 29 April, 11 turtles will be tracked as they migrate from their nesting areas in Costa Rica to south of the Galápagos Islands off Ecuador. The racers are equipped with satellite tags so their locations can be tracked online. The data will provide a nearly real-time, turtle's-eye perspective on the ocean, including measurements of water temperature and depth.

To pick your turtle, visit www.GreatTurtleRace.com.

You could be next

Yes, it *can* happen to you:

If you're a young scientist making inroads in neurobiology research, the next Eppendorf and *Science* Prize for Neurobiology could be yours!

This annual research prize recognizes accomplishments in neurobiology research based on methods of molecular and cell biology. The winner and finalists are selected by a committee of independent scientists, chaired by the Editor-in-Chief of *Science*. Past winners include post-doctoral scholars and assistant professors.

To be eligible, you must be 35 years of age or younger. If you're selected as this year's winner, you will receive $25,000, have your work published in the prestigious journal *Science* and be invited to visit Eppendorf in Hamburg, Germany.

Get recognized!
Deadline for entries:
June 15, 2007

www.eppendorf.com/prize
www.eppendorfscienceprize.org

$25,000 Prize

"I feel incredibly honored to receive this prestigious prize."

2006 Winner
Doris Tsao, Ph.D.
Institute for Brain Research
University of Bremen
Germany

eppendorf & Science PRIZE FOR NEUROBIOLOGY

Triumphs

A CLEAN VICTORY. Kirsten Engel was in the shower when her husband, Scott Saleska, cried out "We won!" after reading about the 2 April ruling by the U.S. Supreme Court that greenhouse gases were pollutants under the Clean Air Act. The win was a very personal one for the University of Arizona, Tucson, faculty members. Saleska, an ecologist, and Engel, an environmental law professor, were key authors on a friend-of-the-court brief that argued in favor of regulating greenhouse gases. After Engel linked up with other lawyers involved in the suit, which was filed by a dozen states and other governmental entities against the U.S. Environmental Protection Agency (EPA), Saleska recruited scientists to help write the brief. The document argued that cutting auto emissions would substantially mitigate climate change. Justice John Paul Stevens cited it when the Supreme Court heard the case last fall, and the 5–4 decision included lengthy references to climate research data mentioned in the brief.

Ironically, Engel and Saleska met each other in 1987 while working at EPA. From there, the pair entered academia, and they've collaborated on a few law review articles since. "Science policy played a role in our coming together," says Saleska.

AWARDS

INNOVATORS. Chemical-sensing polymers that match a dog's ability to sniff out explosives are keeping U.S. soldiers out of harm's way—and have won Massachusetts Institute of Technology (MIT) chemist Timothy Swager (below) this year's $500,000 Lemelson-MIT Prize.

The polymers, which change color when they detect their molecular targets, are the basis for bomb detectors made by an Oklahoma company called Nomadics Inc. U.S. soldiers in Iraq currently analyze people, clothing, and automobiles using the detectors, which are also part of a robotic system for prowling through danger zones. They are among the many contributions that earned Swager one of the country's richest prizes for inventors.

The program also bestowed its first $100,000 prize for sustainability on Dartmouth College chemical engineer Lee Lynd. Over 3 decades, Lynd has created a raft of technologies for turning agricultural wastes and forest trimmings into automotive fuel. He recently co-founded a company, called Mascoma Corp., to commercialize the technology.

MOVERS

TAKING OVER AT NIH. Two acting directors have been named permanent chiefs of their respective institutes at the National Institutes of Health (NIH) in Bethesda, Maryland.

Griffin Rodgers, 52, will head the $1.8 billion National Institute of Diabetes and Digestive and Kidney Diseases, where he has been acting director since Allen Spiegel left last March. A molecular hematologist, Rodgers has spent his career at NIH, where he helped pioneer treatments for sickle cell anemia.

Hematologist Barbara Alving, 60, will direct the $1.1 billion National Center for Research Resources, which she has led in a temporary capacity since 2005. Alving is a former deputy director at the heart institute and head of the Women's Health Initiative.

BIG SHOES. "Wanted: A world-renowned researcher to advise the British Prime Minister on all matters scientific. Knighthood almost guaranteed for good service." The British government has put out a job ad along those lines now that Cambridge University chemist David King, one of the most influential chief scientific advisers in the United Kingdom in recent years, is scheduled to finish his tenure.

King took up the reins in 2000 just before an outbreak of foot-and-mouth disease struck British farms. His advice on slaughtering guidelines is thought to have had a major impact on containing the disease. He contributed to an energy-policy review that controversially recommended a new generation of nuclear power plants. And he's probably best known for proclaiming in 2004 that "climate change is the most severe problem we are facing today—more serious even than the threat of terrorism."

Parliamentarian Ian Gibson, former chair of the House of Commons' Science and Technology Committee, says King "knows his science" and commends him for having "stood up to the American government" on climate change. King will step down by the end of this year.

Three Q's >>

John Mather won a Nobel Prize in physics last year for helping to explain the big bang. Now he's taking on what some would say is an even tougher job. Last week, Mather was named chief scientist in NASA's science office, with the goal of helping his boss, Alan Stern, rescue an imperiled space science program. Mather will split his time between Goddard Space Flight Center in Maryland and NASA headquarters.

Q: Why would a Nobel laureate want this job?
I didn't need to add this to my résumé. But we have an entire planet of people complaining that NASA is not doing the right thing. We need to show that we have a good team and a good plan.

Q: What's your biggest challenge?
Understanding earth science. I've got most of my information until now [from] watching the Weather Channel and Al Gore's movie.

Q: Is your inclination to kill projects or spread the pain?
My instinct is to spread the pain, but experience says that is a bad idea. Then everyone hurts and nothing gets done.

NEWS>>

THIS WEEK

Q&A with Japan's
science adviser
186

Plastic
metals
187
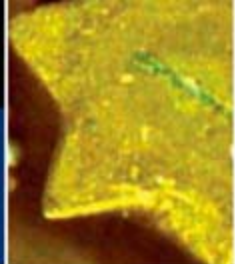
INTELLECTUAL PROPERTY

U.S. Patent Office Casts Doubt on Wisconsin Stem Cell Patents

Opponents of the stem cell patents held by the Wisconsin Alumni Research Foundation (WARF) were delighted last week when the government issued a preliminary ruling rejecting the patents. Critics have long argued that they are far too broad, covering technology that was already in use to derive mouse stem cells and laying claims on all primate embryonic stem (ES) cells in the United States regardless of where they may have been derived. At the same time, patent experts caution that it could take years before the matter is resolved.

The 2 April ruling by the U.S. Patent and Trademark Office (PTO) came in response to a "request" from two public interest groups for a reexamination of three WARF patents awarded in 1998, 2001, and 2006 (*Science*, 21 July 2006, p. 281). The patents assert rights over not only the methodology for cultivating primate ES cells but also, controversially, the cells themselves (*Science*NOW, 3 April, sciencenow. sciencemag.org/cgi/content/full/2007/403/2). Those claims affect the activities of anyone in the United States using human ES cells for either research or commercial purposes. The ruling throws into question patents that reportedly have earned WARF $3.5 million in licensing fees over the past 5 years.

But critics face long odds in battling WARF. Grady Frenchick, a patent lawyer in Madison, Wisconsin, says the PTO initially rejects patents in 90% of reexamination requests, but only 12% of questioned patents are ultimately thrown out. The rest are affirmed in toto or with some modifications. Nonetheless, patent lawyer Cathryn Campbell of San Diego, California, says the WARF decision is more thorough and detailed "than might usually be expected." She also says each of the three patent rejections was signed by a different examiner, suggesting that the conclusions are widely shared in the PTO.

"We're not deluding ourselves that this isn't a tough fight," says WARF Managing Director

Carl Gulbrandsen. If the PTO rules against WARF, WARF will go to the PTO's Board of Patent Appeals and Interferences. If that doesn't work, he says, it's on to the Federal Circuit Court of Appeals in Washington, D.C.

But preliminary as it is, many people are gratified by the patent decision. "Nobody wanted to do anything, but everybody seemed very, very glad that we did," says John M. Simpson of the Foundation for Taxpayer and Consumer Rights in Santa Monica, California, which brought the request last July.

Most scientists doing basic stem cell research in academic or government labs are minimally restricted by WARF's current policies, which require them to pay only $500 for

Road rage. The WARF patents have taken a toll on stem cell researchers.

a batch of Wisconsin cells. But they object to the red tape involved. "Every possible collaboration … is slowed considerably by having to negotiate the WARF Material Transfer Agreements," says George Daley of Harvard Medical School in Boston, Massachusetts. He says if the same rules were applied to mouse cells, "our research would grind to a halt." Martin Pera of the University of Southern California in Los Angeles says that WARF's grip on "basic platform technology critical to the future development" of the

field is bound to impede progress.

For researchers working on commercial applications, WARF makes life much more difficult. WARF spokesperson Andrew Cohn says it's "too complicated" to explain their rates, but Jonathan Auerbach of GlobalStem Inc. in Rockville, Maryland, says he's heard of research licenses costing up to $400,000. Auerbach says many companies have also been put off by WARF's "reach-through" provisions, which call for royalties on any product developed using the cells. He says his company doesn't have to deal with WARF because it has chosen instead to use human embryonal carcinoma cells and human ES cells with abnormal karyotypes that wouldn't be covered by the patents. Mahendra Rao of Invitrogen in Carlsbad, California, says that his company—which is currently negotiating for a WARF license—and others have established outposts outside the United States, where the patents do not apply.

Robert Lanza of Advanced Cell Technology in Worcester, Massachusetts, says his company's WARF license entails "a six-figure fee, plus an annual maintenance charge." On top of that, "whenever a researcher asks us for some ES cells—even ES lines we derived ourselves—we are obligated to pay WARF $5000." Product development is also hobbled. Geron Corp. in Menlo Park, California, has an exclusive license from WARF to develop treatments based on specialized cells grown from the Wisconsin lines, Lanza says, so "we would be sued if we even tried to develop insulin-producing cells to treat diabetes."

Some users are hoping that the widespread complaints could lead WARF to soften its policies further even as the patent reexamination grinds on. In January, for example, WARF lifted the requirement that companies must obtain a license to sponsor human ES cell research at universities. It also eased cell-transfer provisions, lifting fees for transfer of cells among academic and nonprofit researchers. Some critics say both decisions were influenced by the pending patent request. WARF denies this, explaining instead that the decisions are part of "evolving policies … always in favor of increasing access."

–CONSTANCE HOLDEN

ILLUSTRATION: PAT N. LEWIS

FOCUS

Global warming,
round 2
188

A quiet
revolution
196

Shock brigade. Some 800 North Korean soldiers helped erect Pyongyang University.

NORTH KOREA

A Mission to Educate the Elite

SEOUL—In a dramatic new sign that North Korea is emerging from isolation, the country's first international university has announced plans to open its doors in Pyongyang this fall.

Pyongyang University of Science and Technology (PUST) will train select North Korean graduate students in a handful of hard-science disciplines, including computer science and engineering. In addition, founders said last week, the campus will anchor a Silicon Valley–like "industrial cluster" intended to generate jobs and revenue.

One of PUST's central missions is to train future North Korean elite. Another is evangelism. "While the skills to be taught are technical in nature, the spirit underlying this historic venture is unabashedly Christian," its founding president, Chin Kyung Kim, notes on the university's Web site (www.pust.net).

The nascent university is getting a warm reception from scientists involved in efforts to engage the Hermit Kingdom. "PUST is a great experiment for North-South relations," says Dae-Hyun Chung, a physicist who retired from Lawrence Livermore National Laboratory and now works with Roots of Peace, a California nonprofit that aims to remove landmines from Korea's demilitarized zone. To Chung, a Christian university is fitting: A century ago, Christianity was so vibrant in northern Korea, he says, that missionaries called Pyongyang "the Jerusalem of the East."

The idea for PUST came in a surprise overture from North Korea in 2000, a few months after a landmark North-South summit. A decade earlier, Kim had established China's first foreign university: Yanbian University of Science and Technology, in Yanji, the capital of an autonomous Korean enclave in China's Jilin Province, just over the border from North Korea. In March 2001, the North Korean government authorized Kim and his backer, the nonprofit Northeast Asia Foundation for Education and Culture (NAFEC), headquartered in Seoul, to establish PUST in southern Pyongyang. It also granted NAFEC the right to appoint Kim as PUST president and hire faculty of any nationality, as well as a contract to use the land for 50 years.

NAFEC broke ground in June 2002 on a 1-million-square-meter plot that had belonged to the People's Army in Pyongyang's Nak Lak district, on the bank of the Taedong River. Construction began in earnest in April 2004. That summer, workers—a few of the 800 young soldiers on loan to the project—unearthed part of a bell tower belonging to a 19th century church dedicated to Robert Jermain Thomas, a Welsh Protestant missionary killed aboard his ship on the Taedong in 1866.

NAFEC's fundraising faltered, however, and construction halted in fall 2004. The group intensified its Monday evening prayers and broadened its money hunt, getting critical assistance from a U.S. ally: the former president of Rice University, Malcolm Gillis, a well-connected friend of the elder George Bush and one of three co-chairs of a committee overseeing PUST's establishment. "He made a huge difference," says Chan-Mo Park, president of Pohang University of Science and Technology (POSTECH), another co-chair. South Korea's unification ministry also quietly handed PUST a $1 million grant—more than it has awarded to any other North-South science cooperation project. This helped the school complete its initial $20 million construction push.

At the outset, PUST will offer master's and Ph.D. programs in areas including computing, electronics, and agricultural engineering, as well as an MBA program. North Korea's education ministry will propose qualified students, from which PUST will handpick the inaugural class of 150. It is now seeking 45 faculty members. Gillis and other supporters are continuing to stump for a targeted $150 million endowment to cover PUST operations, which in the first year will cost $4 million. Undergraduate programs will be added later, officials say. PUST, at full strength, aims to have 250 faculty members, 600 grad students, and 2000 undergrads.

PUST hopes to establish research links and exchanges with North Korea's top institutions and with universities abroad. "It is a very positive sign," says Stuart Thorson, a political scientist at Syracuse University in New York who leads a computer science collaboration between Syracuse and Kimchaek University of Technology in Pyongyang. "Key to success will be achieving on-the-ground involvement of international faculty in PUST's teaching and research."

Some observers remain cautious, suggesting that the North Korean military could use the project to acquire weapons technology or might simply commandeer the campus after completion. A more probable risk is that trouble in the ongoing nuclear talks could cause delays. At the moment, however, signs are auspicious. Park, who plans to teach at PUST after his 4-year POSTECH term ends in August, visited Pyongyang last month as part of a PUST delegation. "The atmosphere was friendly," he says. "The tension was gone." The Monday prayer group continues, just in case. **–RICHARD STONE**

Study Finds Foreign High-Tech Workers Earn Less

Many U.S. companies say they hire foreign scientists and engineers because of a shortage of qualified native-born workers. But a new salary study bolsters the claim of some analysts that a strong reason may be to hold down wages.

The study, by B. Lindsay Lowell and Johanna Avato of Georgetown University in Washington, D.C., shows that science, technology, engineering, and mathematics (STEM) workers holding an H-1B—a temporary visa granted to skilled foreign workers—earn 5% less than natives employed in similar positions with similar skills and experience earn. It also shows that H-1B visa holders who don't job-hop make 11% less than natives and that those who enter the workforce after graduating from a U.S. university earn 16% less.

There is one group of foreigners who do not seem handicapped by their H-1B visa status, however: Those hired directly from overseas—45% of the total—make 14% more than native workers. The study, presented last month before the Population Association of America, uses data collected by

the National Science Foundation as part of a 2003 National Survey of College Graduates.

These findings could influence pending legislation affecting a program that every year admits 65,000 foreign nationals into the U.S. workforce. Business groups want Congress to greatly increase—or, better still, eliminate—the existing ceiling on H-1B visas, arguing that it hurts U.S. competitiveness. The workers, many from India and China, are in such high demand that this month, the government received applications for more than twice the number of slots available next year—on the

Fair market? Overall, H-1B visa holders earn 5% less than native-born U.S. workers holding similar jobs. But the difference varies by category of worker.

very first day the applications could be filed.

Lowell speculates that foreign workers are paid less because they are often compelled to remain with the same employer to get permanent residency within the 6 years of stay allowed by their visas. Lowell says this "de facto bondage"—the residency process, which can take years, starts anew if they change jobs—has the effect of depressing salaries not just for foreign workers but for natives as well.

One solution, Lowell says, is to grant permanent residency to foreign workers right off the bat, or at least waive the requirement that applicants be sponsored by their employer. Indeed, several bills would grant automatic permanent residency to foreign students graduating from U.S. institutions with advanced STEM degrees (*Science*, 14 April 2006, p. 177).

Opponents of high-tech immigration, however, say that the salary differential between H-1B visa holders and natives argues for ending the H-1B program. "Either these foreign temporary workers are not 'the best and the brightest,' or companies are hiring them to hold down starting wages—or both," says Jack Martin of the Federation for American Immigration Reform in Washington, D.C.

–YUDHIJIT BHATTACHARJEE

NSF to Revisit Cost-Sharing Policies

Cost sharing has long been a requirement for many types of competitive grants at the National Science Foundation (NSF). In 2001, for example, institutions pledged more than half a billion dollars to supplement some 3300 NSF-funded projects on their campuses. But despite its value in leveraging federal dollars, cost sharing can also give wealthier institutions an unfair advantage in vying for an award. So in October 2004, NSF decided to eliminate the provision from future program announcements.

Now NSF's oversight body, the National Science Board, wants to take another look at the issue. Some board members worry that local and state governments, industry, and other nonfederal research partners may lose interest in research collaborations if they don't have a financial stake in the project. "The original idea was to bring in more money, but I think cost sharing is really more about building partnerships," says Kelvin

Droegemeier, a meteorology professor at the University of Oklahoma in Norman who volunteered to lead the board's reexamination. "The institutional buy-in is an important element, and I wonder if the board went too far [in 2004] when we eliminated it."

The decision to reopen a long-running debate disturbs some university administrators, who note that federal funding already falls far short of paying for the full cost of academic research. "We had been urging NSF to end [cost sharing] for many years because of our concern about how it was being used in the evaluation process," explains Anthony DeCrappeo of the Council on Government Relations, a Washington, D.C.–based association of research universities.

DeCrappeo says grant applicants often suspected a subtle bias from reviewers and program managers in favor of proposals with large institutional commitments. Schools were confused about which programs

required cost sharing, he adds. Finally, institutions at times came up with their share by diverting money from existing research activities. Universities spent $8 billion a year on academic research in 2005—more than either companies or state governments, he notes, and only some of which represents federal reimbursement for the cost of supporting research on campus—"and there's no reason to have additional matching requirements."

Droegemeier says that the board hopes to collect data on the impact of cost sharing across different NSF programs. He and others are especially concerned about no longer requiring state legislatures to support the Experimental Program to Stimulate Competitive Research operating in 27 states and territories that receive relatively small amounts of NSF funding. "We want to get community feedback," he says. "But something tells me that [eliminating cost sharing] is not the best way to go." **–JEFFREY MERVIS**

GENETICS

Mysterious, Widespread Obesity Gene Found Through Diabetes Study

The role that obesity plays in diabetes, cancer, and other diseases makes our expanding waistlines one of today's most pressing health problems. Now, on the genetics front, researchers have nabbed a coveted prize: the first clear-cut evidence for a common gene that helps explain why some people get fat and others stay trim. The British team, led by Andrew Hattersley of Peninsula Medical School in Exeter and Mark McCarthy of Oxford University, doesn't know what this gene, called *FTO*, does. But adults, and even children, with two copies of a particular *FTO* variant weighed on average 3 kilograms more than people lacking the variant, the researchers report in a paper published online by *Science* this week (www.sciencemag.org/cgi/content/abstract/1141634).

Although twin studies have suggested that obesity has a genetic component, some earlier reports of common obesity genes, including a paper in *Science* last year (14 April 2006, p. 279), have proved controversial. But this new work, which involved nearly 39,000 people, is solid, says Francis Collins, director of the U.S. National Human Genome Research Institute in Bethesda, Maryland. "There's no question that this is correct."

The U.K. team first found the gene in type 2 diabetes patients participating in a multidisease study sponsored by the Wellcome Trust, the U.K. biomedical charity. Timothy Frayling in Hattersley's lab and his co-workers first analyzed the genomes of 1924 diabetic and 2938 nondiabetic individuals, looking for which of nearly 500,000 genetic markers were more common in those with diabetes. Those markers helped them home in on a variant, called a single-nucleotide polymorphism, in the *FTO* gene. The gene, located on chromosome 16, was a surprise: Whereas other known diabetes genes predominantly control insulin production, *FTO* proved to be associated with body mass index, or BMI (weight divided by height squared)—suggesting that it might control weight in more than just people with diabetes.

To find out, 41 collaborators looked for the *FTO* mutation in DNA samples from "literally every single study we could," says Hattersley, including another two diabetes populations, nine cohorts of white European adults, and two studies of European children. In every one, the *FTO* mutation was associated with BMI. Overall, about 16% of white adults and children carry two copies of this variant. They are 1.67 times more likely than those lacking any copies to be obese, the group reports.

The researchers don't know what *FTO* does. But because *FTO* may lead to a new pathway for controlling weight, "we'll have people racing to understand" the gene's function, says obesity researcher Jeffrey Flier of Harvard Medical School in Boston. Those studies should help unravel the basic biology of obesity.

The paper also underscores the importance of tracking down common disease genes in as many groups of people as possible. In the past 2 years, researchers have reported finding common disease genes for age-related macular degeneration, diabetes, and—earlier this month—prostate cancer. However, the finding of another obesity gene, *INSIG2*, published last year in *Science*, has held up in only five of nine populations, says co-author Michael Christman of Boston University. The case for *FTO*'s involvement is strengthened by the fact that other obesity gene hunters are finding the *FTO* polymorphism as well. "[There's] very strong evidence that it's a gene that affects body weight," says human geneticist David Altshuler of the Broad Institute in Cambridge, Massachusetts. "That's very exciting."

–JOCELYN KAISER

Flab factor. A genetic variant appears to affect some people's body weight.

CREDIT: JEFF J. MITCHELL/GETTY IMAGES

SCIENCE POLICY

Japan Picks Up the 'Innovation' Mantra

TOKYO—Kiyoshi Kurokawa, science adviser to Japan's prime minister since last fall, doesn't mince words when it comes to talking about what's best for Japan's research and development efforts. "First you have to reform the leading universities," he says.

Kurokawa, 70, was offered the job when a phone call "came completely out of the blue" from just-elected Prime Minister Shinzo Abe's office late last September. It was the first time a Japanese prime minister had appointed a science adviser. Kurokawa suspects he caught Abe's attention with his outspoken opinions given while serving on the governmental Council for Science and Technology Policy and as president of the Science Council of Japan. The position is not permanent and could disappear if Abe fails to lead the Liberal Democratic Party to success in elections later this year.

Kurokawa led the drafting of "Innovation 25," Abe's vision of how science and technology can contribute to Japan's economic growth out to 2025. Kurokawa laughs about "innovation" being in the title of so many recent science policy manifestos. But he firmly believes in the recommendations, which include making energy and the environment drivers for economic growth, radically increasing funding for education, and reforming Japan's universities.

University reform is a pet topic for Kurokawa, who rose to be a professor of medicine at the University of California, Los Angeles, before returning to Japan where, after a stint at the University of Tokyo, he became dean of the School of Medicine of Tokai University in Hiratsuka, Kanagawa Prefecture. Below are his edited comments from an interview with *Science*.

–DENNIS NORMILE

On innovation:

The innovation Abe is talking about is not just technological innovation, but social innovation and also nurturing innovative people. Japanese society has to become more conducive to innovation and provide opportunities for risk-taking, adventurous people. It's fine to invest in science and technology. That provides the seeds for [economic] value. But in this globalized age, you really have to compete and deliver the seeds of scientific discovery to the marketplace. That requires social encouragement of entrepreneurial activities.

On shifting government spending from public works to human resources:

[The Innovation 25 plan] is a sort of vision statement by the government, and each ministry will be asked to follow this road map. The overall annual budget should have certain objectives. But it is very hard to change [priorities], because each ministry has its own [interests] and their budget remains more or less the same from year to year.

We could shift public spending more toward human resources rather than infrastructure. But because of the political decision-making process, you have to raise public awareness so that any politician [endorsing a shift] will be supported. As science adviser to the prime minister, I'll try to [do that].

On reforming Japan's universities:

At the leading universities, you have to choose when taking the entrance exam [which academic department] you are heading toward. Even within a school of engineering, you have to choose, say, electrical engineering. This means that even by grade 10, students' core studies are shifting depending on whether they want to go into the natural sciences or social sciences or arts and humanities. Why does it have to be this way? Let high school students study whatever they are interested in and get universities to allow more flexible choices. Right now in Japanese society, if it so happens that at age 18 you didn't study [and failed to enter university], there's no second chance. Universities should have more flexible entrance policies.

For another example, [internationalize the universities] by aiming for 30% of undergraduates to be foreigners. Give them scholarships if need be. The impact of their presence would be to change the mindset of Japanese. There has been talk about Japan becoming a very attractive place for foreign researchers to come for graduate study. Let's start at an earlier age.

Finally, you have to reform the Japanese hierarchical academic system [in which junior researchers work under department chairs]. That destroys the creativity and independence of younger professors. Under this inbred system, you're just nurturing cloned professors.

On the scientific community's responsibility to the public:

People have higher expectations for contributions from the science community because their money is spent on research and development. The public is more informed, and they want more return on their investment—and that's natural. …

The science community should be accountable for their policy recommendations. Whether the science community becomes trusted by the public depends on doing that. So I think transparency and engagement with the public is very important.

Radical overhaul. Kiyoshi Kurokawa, science adviser to Japan's prime minister, wants more money for education, but only if universities reform.
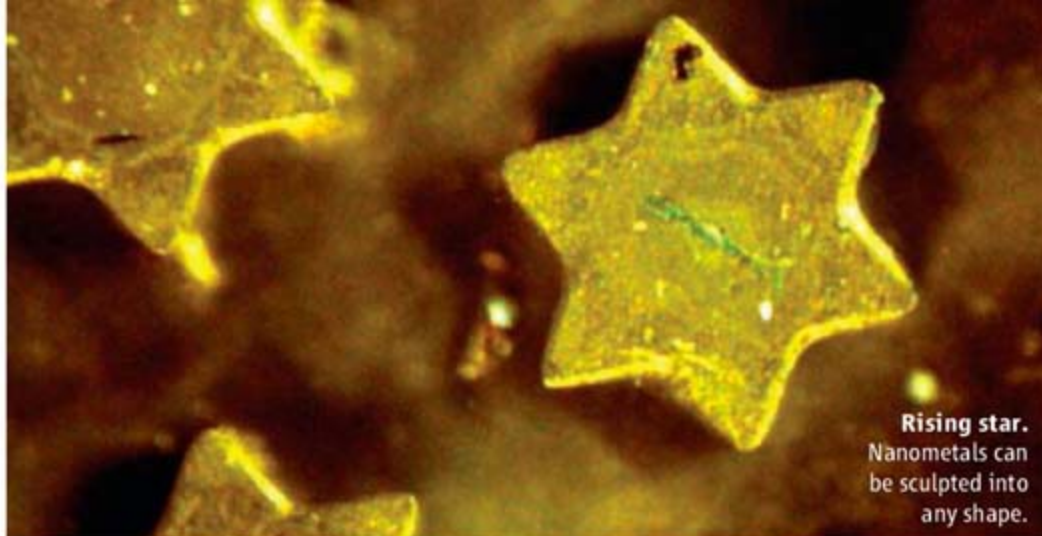
Rising star. Nanometals can be sculpted into any shape.

MATERIALS SCIENCE

Chemists Mold Metal Objects From Plastic 'Nanoputty'

Blacksmiths have molded metals for thousands of years by melting them at ultrahigh temperatures. Now, much like potters transforming clay into ceramics, a group of chemists has found a way to assemble tiny metal particles into a substance that can be shaped and fired—at little more than room temperature. The process creates objects composed of either a single metal or alloys of multiple metals, which could make them well-suited for a raft of applications including catalysis and optics.

The new work, described on page 261, is drawing high praise. "It's a very nice way to mold particles into whatever shape you want," says Chuan-Jian Zhong, a chemist at Binghamton University in New York, who describes the work as "excellent."

Nanoparticles are the focus of intense research because their tiny size lends them unique electrical, chemical, and optical properties. But when researchers try to join them into assemblies, the particles typically create rigid crystals that can't be reshaped. So Bartosz Grzybowski, a chemist at Northwestern University in Evanston, Illinois, set out to give nanoparticle assemblies a little flexibility. That required striking a very delicate balance. If the nanoparticles bond too readily to each other, each particle winds up linked to all its neighbors, resulting in a tightly knotted ball. But if too few connections are made between particles, the assembly doesn't grow.

Grzybowski and his colleagues started by creating linkers consisting of long hydrocarbon chains sporting thiol groups at each end that readily bind to metal particles. In the middle of the linkers, they placed azobenzene groups that change their conformation when exposed to ultraviolet light—in this case, switching the linkers from oil-friendly hydrophobic compounds to water-friendly hydrophilic ones.

The researchers dissolved the linkers in a mixture of an organic solvent and soaplike surfactant and added metal nanoparticles coated with organic compounds abbreviated DDA. As the nanoparticles dispersed through the solution, thiol groups on one end of the linkers displaced weakly binding DDA molecules, glomming onto individual nanoparticles. At this stage, each metal particle was coated by DDA molecules and a few linkers, and surfactants surrounded the linkers' free thiol groups so they did not "see" any of the metal nanoparticles floating nearby. When Grzybowski's team flipped on the UV light, the linkers became hydrophilic and migrated toward one another in the hydrophobic organic solvent. The free thiol groups latched onto nanoparticles on neighboring linkers, creating growing webs of particles.

The Northwestern team didn't want all these webs to unite, however, because that would lead to a messy precipitate. After some trial and error, they found that if they added just the right amount of nanoparticles, a large number of spherical webs would form, but the particle feedstock ran out before they joined up. Together, these "supraspheres" formed a kind of waxy paste the consistency of putty, which could be molded to form essentially any shape from spheres to gears. Moreover, when they fired their shapes at a modest 50°C, the heat drove off the organics and welded the neighboring nanoparticles together, creating a continuous and porous metal network.

The Northwestern researchers have already shown that their newly fired metals are electrically conductive. Now they are testing their optical and catalytic properties. If those turn out well, moldable nanometals may end up in everything from catalytic membranes for fuel cells to novel chemical sensors.

–ROBERT F. SERVICE

SCIENCE**SCOPE**

Congress Probing Enviro Institute

A deadline looms next week for David Schwartz to respond to congressional questions about his office's spending and his other activities as director of the National Institute of Environmental Health Sciences (NIEHS).

Henry Waxman (D–CA), chair of the House Committee on Oversight and Government Reform, and committee member Dennis Kucinich (D–OH) wrote in a 30 March letter that Kucinich is following up a January inquiry regarding Schwartz's controversial efforts to revamp NIEHS's journal *Environmental Health Perspectives*. Last year, Schwartz scrapped a proposal to privatize the journal, but critics still question his plans to cut costs.

But now, spurred by what it calls new information from "multiple sources," the committee also wants documents on Schwartz's activities as director, including his office's budget and any consulting or travel he's done for outside organizations under the National Institutes of Health's strict new ethics rules. A committee spokesperson declined to discuss the new information it had received on Schwartz, and an NIEHS spokesperson says it is "putting [its] responses together."

–JOCELYN KAISER

Judge Takes Ax to Forest Plan Changes

Federal agencies misrepresented and buried the views of dissenting scientists when they decided to make logging easier in the Pacific Northwest, a U.S. district judge ruled last week. In his decision, Judge Ricardo Martinez tossed out the agencies' changes to the Northwest Forest Plan, which puts tight constraints on old-growth logging (*Science*, 29 July 2005, p. 688).

In 2003, the U.S. Forest Service (USFS) and other agencies proposed to amend the way that watersheds are evaluated before logging projects are approved. A number of eminent scientists noted their concern that the amendment would "remove or weaken several key conservation provisions for aquatic species." Martinez ruled that these concerns were not prominently mentioned in the draft Environmental Impact Statement, as required by law, and were misrepresented in a summary of comments. The agencies were "trying to spin what was going on," says Patti Goldman, an attorney with Earthjustice. The agencies now have until late June to decide whether they will appeal the ruling.

–ERIK STOKSTAD

Global Warming Is Changing the World

An international climate assessment finds for the first time that humans are altering their world and the life in it by altering climate; looking ahead, global warming's impacts will only worsen

IN EARLY FEBRUARY, THE UNITED NATIONS–sponsored Intergovernmental Panel on Climate Change (IPCC) declared in no uncertain terms that the world is warming and that humans are mostly to blame. Last week, another IPCC working group reported for the first time that humans—through the greenhouse gases we spew into the atmosphere and the resulting climate change—are behind many of the physical and biological changes that media accounts have already associated with global warming. Receding glaciers, early-blooming trees, bleached corals, acidifying oceans, killer heat waves, and butterflies retreating up mountainsides are likely all ultimately responses to the atmosphere's growing burden of greenhouse gases. "Climate change is being felt where people live and by many species," says geoscientist Michael Oppenheimer of Princeton University, a lead author of the report. "Some changes are making life harder to cope with for people and other species."

The latest IPCC report (www.ipcc.ch/SPM6avr07.pdf) sees a bleak future if we humans persist in our ways. The climate impacts, mostly negative, would fall hardest on the poor, developing countries, and flora and fauna—that is, on those least capable of adapting to change. Even the modest climate changes expected in the next few decades will begin to decrease crop productivity at low latitudes, where drying will be concentrated. At the same time, disease and death from heat waves, floods, and drought would increase. Toward midcentury, up to 30% of species would be at increasing risk of extinction.

"This stark and succinct assessment of the future … is certainly troubling," wrote economist and coordinating lead author Gary Yohe of Wesleyan University in Middletown, Connecticut, in an e-mail message from the final meeting of the IPCC working group in Brussels, Belgium. It is now obvious, he says, that even if greenhouse gas emissions are immediately reduced, changes are inevitable. Humans will have to adapt, if we can.

Toning down the message

The working group's report had a difficult coming-out party on 6 April. Like the reports from the two other IPCC working groups (WGI—see *Science*, 9 February, p. 754—and WGIII, due out on 4 May), Working Group II's involved a couple of hundred scientist authors from all six continents analyzing and synthesizing the literature over several years. Reviews by hundreds of experts and governments generated thousands of comments. Twenty chapters totaling 700 printed pages led to a Technical Summary of 80 to 100 pages and a Summary for Policymakers (SPM) of 23 pages. Then came the hard part: the 4-day plenary session in Brussels, which brought together scientists and representatives of 120 governments. There, unanimity among governments is required on every word in the SPM, ostensibly to ensure that the phrasing clearly and faithfully reflects the reviewed science of the chapters.

This time, there were "bigger bumps than normal," says climate scientist Stephen Schneider of Stanford University in Palo Alto, California, a coordinating lead author. "It was longer and more painful than usual," Oppenheimer agrees. Especially as the deadline approached early Friday morning, a few countries—attendees mention coal-rich China and oil-rich Saudi Arabia most often—insisted on substantial changes. Sometimes, the softening of the summary could be taken as a technical adjustment. For example, the SPM draft's "20 to 30% [of] species at increasingly high risk of extinction" as the world warms 1° or 2°C became "Up to 30% of species at increasing risk of extinction."

Perhaps the most substantial loss from the draft SPM was in the tables. The plenary session eliminated parts of a table that would

Drought will return to southwest North America.

Wetter's better? Warmer and wetter high latitudes will yield more crops but also more flooding.

> "For the first time, we concluded **anthropogenic warming has had an influence on many physical and biological systems.**"
>
> —Cynthia Rosenzweig, Goddard Institute for Space Studies

Winters in Northern Europe will be less severe.

Arctic permafrost will thaw.

The Mediterranean region will dry out.

Savanna will replace tropical forests.

Rising sea level will increase coastal flooding.

IPCC's Projected Impacts

Mountain glaciers will disappear.

Most corals will suffer major declines.

MAP PHOTO CREDITS (LEFT TO RIGHT): GEORGE E. MARSH/AP/NOAA; KAREL NAVARRO/AP; JOHN MAIER JR./LPI; STEVE PARKIN/AFP/GETTY IMAGES; JUAN MANUEL SERRANO/AP; ROGER TIDMAN/CORBIS; FARJANA K. GODHULY/AFP/GETTY IMAGES; OVE HOEGH-GULDBERG/AP

have allowed a reader to estimate when in this century the various projected impacts might arrive. Also dropped was an entire table that laid out quantified impacts—such as annual bleaching of the Great Barrier Reef in the relatively near term—in an easily accessible, region-by-region format.

Toning-down aside, "it's still a decent report," says Schneider. "There are no key science points that didn't come through in the SPM," says ecologist Christopher Field of Stanford, a coordinating lead author. And all of the losses from the draft SPM are still available in the Technical Summary and the underlying chapters for the determined reader. However, anyone reading the SPM "should understand that the findings are stated very conservatively," says Field.

Impacts, present and future

Conservative though it may be, the report holds one major first. "For the first time, we concluded anthropogenic warming has had an influence on many physical and biological systems," says impacts analyst and coordinating lead author Cynthia Rosenzweig of NASA's Goddard Institute for Space Studies in New York City. Media coverage of weird weather and its effects had come to imply that global warming was affecting things both living and inanimate, and individual studies pointed that way too, but no official body had given the link its imprimatur.

To make it official, IPCC authors considered 29,000 series of observations from 75 studies. Of those series, 89% showed changes—glaciers receding or plants blooming earlier, for example—consistent with a response to warming. Those responses so often fell where greenhouse warming has been greatest that it's "very unlikely" the changes were due to natural variability of climate or of the physical or biological system involved. "It's clear it's not all about future impacts," says Field. As an example, he cites the decline of more than 20% in snowmelt since 1950 as the U.S. Pacific Northwest has warmed. That puts a squeeze on everything from hydroelectric dams to salmon.

Like the ongoing effects of global warming, future impacts will vary greatly from region to region. Perhaps the most striking example is shifting precipitation. WGII authors started with WGI's model-based prediction of increasing dryness at low latitudes (the U.S. Southwest and northern Mexico; the Caribbean region, including northeast Brazil; and all around the Mediterranean) and increasing wetness at high latitudes (northern North America and northern Eurasia). They then drew on published studies of the effects of climate change on crops.

The results of a meta-analysis of 70 modeling studies "are compelling," says geographer William Easterling of Pennsylvania State University in State College, a coordinating lead author. "It's become very clear that in high latitudes, a warming of 1° to 3°C is beneficial for the major cereals—wheat, corn, and rice. At

the same time, in low latitudes, even a little warming—1°C—results in an almost immediate decrease in yield." In the north, the added water accompanying warming boosts yields, but toward the equator, the added heat is too much for the plants. But "you can't warm the mid-latitudes forever without getting some negative response," says Easterling. "After a 3°C warming, you get this consistent downturn in cereal yield" even at higher latitudes. A 3°C warming is possible globally late in the century if nothing is done about emissions.

Other global warming impacts are even more localized. As glaciers melt in the next few decades in places such as the Andes and Himalayas, flooding and rock avalanches will increase at first. Then, as the glaciers continue to recede toward oblivion, water supplies will decrease. Sea-level rise from

More ominous is the report's discussion of potentially large sea-level rise. The main statement is low-key: "There is medium confidence that at least partial deglaciation of the Greenland ice sheet, and possibly the West Antarctic ice sheet, would occur over a period of time ranging from centuries to millennia for a global average temperature increase of 1–4°C (relative to 1990–2000), causing a contribution to sea level rise of 4–6 m or more."

Four to 6 meters of sea-level rise would be globally catastrophic. New Orleans, south Florida, much of Bangladesh, and many major coastal cities would be inundated. Centuries to millennia might seem like plenty of time to deal with this still-uncertain prospect, but the "1–4°C" is a tip-off. Combine that with the table of greenhouse gas–emission scenarios dropped from the SPM, and it is evi-

culation doesn't even include many non-quantifiable impacts, such as ecosystem losses and the conflicts resulting from climate refugees, that could double damage costs. The SPM's bottom line: "The net damage costs of climate change are likely to be significant."

Economists are "virtually certain," however, that whatever the global climate costs prove to be, not everyone will bear them equally. Some people will be exposed to more climate change than others. Some will be more sensitive to it. Some will be less able to adapt to it. And some will suffer on all three accounts. These people might live in countries that lie in low latitudes where drying will predominate. Their economies are likely based largely on agriculture that is susceptible to drought. And they are more likely to be developing countries without the wealth needed to adapt to climate change, say, by building irrigation systems.

Because such happenstances of geography, climate, and economics make some groups particularly vulnerable, Yohe says, "climate change will impede progress toward meeting Millennium Development Goals"—eight U.N.-sponsored goals, which include eradicating extreme poverty and hunger and ensuring environmental sustainability. "If you don't do something about climate, you're swimming upstream" trying to meet these goals across the world. Fortunately, says Yohe, many of the steps that would help communities adapt to climate change would also help meet the U.N. goals.

Although the report emphasizes the vulnerability of poorer, developing countries, it foresees no real winners. Every population has vulnerable segments, Oppenheimer points out. In the European heat wave of 2003 that killed perhaps 30,000, it was the elderly. When Hurricane Katrina hit New Orleans, Louisiana, killing 700, it was the poor. Adaptation—building levees in the case of New Orleans—has not worked out all that well so far.

Projected Patterns of Precipitation Changes

multimodel   A1B   DJF   multimodel   A1B   JJA

%   −20 −10 −5   5   10 20

Some of both. Global warming will bring more precipitation (bluish) to high latitudes in both winter (*left*) and summer (*right*) and less precipitation (reddish) to low latitudes.

melting glaciers and ice sheets would flood low-lying coastal areas, threatening tens of millions of people living on the megadeltas of Africa and Asia, such as the Nile and Brahmaputra. Coral lives near its upper limits of temperature, so even modest warming is projected to lead to more frequent bleaching events and widespread mortality. Extreme heat waves would become more frequent and more deadly for people. Warming and drying would encourage forest pests, diseases, and fire, hitting forests harder as larger areas are burned. The IPCC list goes on and on.

The report also briefly considers potentially catastrophic climate events. WGI had already found that in this century, the great "conveyor belt" of currents carrying warm water into the chilly far North Atlantic will only slow, not collapse. So Western Europe isn't about to freeze over. In fact, it would warm under the strengthening greenhouse. But WGII still sees likely North Atlantic–wide effects including lower seawater oxygen and changes in fisheries.

dent that a 1°C warming would in all likelihood arrive by mid-century, assuming no action to cut emissions. A 3°C warming could be here by the end of the century. Although the sluggish ice sheets might not respond completely to that warming for centuries or millennia, before the century is up, the world could be committed to inundation of its low-lying coastal regions.

The world loses

So what's the bottom line? WGII did that calculation too. According to the SPM, "Global mean losses could be 1–5% [of] Gross Domestic Product (GDP) for 4°C of warming." That's a range from significant but bearable to truly burdensome. "There's too much uncertainty in that calculation" to take it too seriously, Yohe says. That's because it is a messy computation involving assumptions about all sorts of factors: how sensitive the climate really is to added greenhouse gases; what people alive today owe to future generations; how to balance the needs of greenhouse gas emitters and climate victims. And the cal-

And no one region seems exempt. In a paper published online by *Science* on 5 April (www.sciencemag.org/cgi/content/abstract/ 1139601), climate modeler Richard Seager of Lamont-Doherty Earth Observatory in Palisades, New York, and his colleagues look at 19 global climate models run for the IPCC. They expect the dryness of the 1930s Dust Bowl to return to the American Southwest by midcentury, for good. If the models are right, the western drought of the past decade is only the beginning. If the world's biggest emitter of greenhouse gases needed some prodding to take action on global warming, this could be it.

–RICHARD A. KERR

IMMUNOLOGY

The Education Of T Cells

New research on how T cells learn to home in on their targets could lead to selective treatments that boost or dampen immune responses in specific tissues of the body

Psst, here's the plan. When a dendritic cell (*top*) embraces a T cell (*bottom*), it activates it and instructs it where to migrate.

Almost 3 decades ago, a team of immunologists made an intriguing observation. They collected white blood cells called lymphocytes from lymphatic fluid (lymph) that drained the skin or the gut of a healthy sheep, labeled those lymphocytes, and injected them back into the same sheep's bloodstream. To their surprise, the injected cells didn't patrol the whole body: Cells from the skin region returned mostly to the skin, whereas those from the intestine homed mostly back to the gut.

T cells, the infection-fighting immune cells born in the thymus, were thought to cruise the entire body via the bloodstream and the lymphatic circulation, stopping where they spotted signs of trouble. So how did those sheep T cells know to navigate to and patrol a particular tissue? The question matters because immunologists hope to battle tumors or autoimmune diseases by controlling the cellular immune response in one organ, while leaving the immune system alone elsewhere.

The first clues to an answer came from Eugene Butcher and Irving Weissman of the Stanford University School of Medicine in Palo Alto, California. In the 1980s, they showed that certain squads of T cells can distinguish between tiny blood vessels near the skin or near the intestine. Then Butcher's team and others identified dozens of cell-surface receptors and soluble signaling chemicals called chemokines that

helped those T cells penetrate and patrol particular tissues. In the 1990s, Butcher and other biologists uncovered a molecular code—the unique combination of receptors and chemokines—that directed T cells to, say, the skin or the gut. But one crucial mystery remained: How does a newborn T cell, fresh from the thymus, become programmed, or educated, to express the combination of receptors that will let them home to a particular tissue? "It's a fundamentally important problem in cellular immunology," says Jeffrey Frelinger of the University of North Carolina, Chapel Hill.

Over the past 5 years, researchers have begun to crack that mystery. The most recent work, which shows how immune sentinels called dendritic cells instruct T cells where to go, is revealing a layer of intelligence in the body's immune surveillance mechanisms that had gone undetected, say Frelinger and other immunologists. Ultimately, physicians hope to use the emerging understanding of T cell targeting to develop immune-modulating compounds more specific than today's drugs, which for the most part are blunt instruments that can cause serious side effects. Drugs that direct T cells to specific sites could battle tumors, improve vaccines, or ease autoimmune diseases. "One can conceivably generate drugs that interfere with organ-specific [T cell] recruitment without paralyzing immune defenses everywhere else," says immunologist Ulrich von Andrian of Harvard Medical School (HMS) in Boston.

T cells on patrol

When tissue is infected by a foreign agent, its first line of defense is inflammation, the nonspecific response involving pain, redness, heat, and swelling. Then, over several days, the immune system activates squads of T cell clones, lines of cells each of which can latch onto a single bit of pathogen on an infected cell. T cells then neutralize the threat, call for backup from other immune cells, or both.

T cell activation begins when dendritic cells, octopuslike cells that roam the body's tissues, spot infection and chew up infected cells to obtain antigen—a small piece of a pathogen or tumor that can trigger an immune response. Dendritic cells then travel through the lymphatic ducts to the nearest lymph node, spongelike sacs that serve as regional field stations for the immune system. There the dendritic cells encounter many so-called naïve T cells but only activate for battle the ones bearing receptors that recognize the antigen they carry. The newly vigilant T cells multiply into an army of clones known as effector T cells that can fight infected or rogue cells.

The effector T cells then move from the lymph nodes through lymphatic vessels to the bloodstream, where they circulate throughout the entire body. But to fight pathogens, they need to find the site of the infection. Immunologists believe that some effector T cells stop in any tissue or organ where there are signs of trouble, or inflammation. But Butcher and others have long concentrated on the more specialized T cells that can home back from the bloodstream to a particular tissue, such as skin or gut.

By the early 2000s, Butcher and others had uncovered a clever addressing system that targets those tissue-specific T cells to the correct home. These T cells use a four-step process to exit the bloodstream across the walls of tiny veins called high endothelial venules. Each of the four steps requires either matching pairs of Velcro-like receptors on T cells and the venule walls, or matching pairs of other T cell receptors and chemoattractants, small molecules that make up a tissue's unique chemical scent. If the four correct pairs of receptors and chemoattractants are present in the right combination, the T cell recognizes that it's in the correct tissue, then squeezes through the venule wall to the tissue beyond. Today, Butcher says, the field is starting to ask how a naïve T cell learns to express the correct combination of homing receptors for the gut, skin, or other tissues—a process called T cell education, or imprinting.

Before immunologists could find out how T cells undergo such imprinting, they had to make sure it really happened in living animals and that the cells were not born "pre-committed to homing to gut or skin or joints," Butcher says. Butcher and Daniel Campbell, now at the University of Washington, Seattle, did that in 2002. They injected mice with millions of identical, fluorescently labeled mouse T cells, all of which had been genetically engineered to recognize an egg-white protein. They immunized the mice with that egg-white protein, then 2 days later, surgically removed lymph nodes and other lymphoid tissue from the gut and the skin. Inside all the lymphoid tissue they examined, the quiescent T cells were being activated into effector T cells that were ready to battle the foreign protein. But T cells found in the gut lymph nodes produced receptors that would help them find their way to the gut itself once they had reentered the bloodstream from the nodes, whereas otherwise identical T cells from the skin lymph nodes produced receptors that would direct them to skin, the researchers reported in the *Journal of Experimental Medicine*. "Where you get stimulated determines which homing receptors are expressed," Butcher explains.

What happens within a tissue's lymph node to program a T cell to migrate from the bloodstream to that tissue? Von Andrian suspected that dendritic cells teach T cells to home to the tissue where those foreign bits are found. That's because dendritic cells are on the scene in lymph nodes, embracing and helping activate the T cells.

Von Andrian's team purified dendritic cells from lymphoid tissue (lymph nodes or other specialized immune tissue) from three parts of the body: spleen (a central lymphoid organ), skin, and intestine. They incubated each tissue-specific type of dendritic cell in separate petri dishes with naïve T cells. After 5 days, T cells were ready to do battle with pathogens. But in a test-tube experiment, only T cells exposed to dendritic cells from the Peyer's patch, lymphoid tissue in the intestinal wall, migrated toward a gut chemokine.

Then, to see whether the same thing happened in living animals, the researchers injected mice with fluorescent T cells that had been stimulated by one of the three types of dendritic cells. T cells ended up mostly in the gut when they'd been activated by dendritic cells from gut lymphoid tissue, but not when they'd been activated by dendritic cells from skin lymph nodes, the researchers reported in 2003 in *Nature*. The same year, immunologist William Agace's

team at Lund University in Sweden reported that dendritic cells from mesenteric lymph nodes, another immune site in the gut, also educate T cells they touch to home in on the intestines. Together, the results mean that "antigen-presenting cells from different lymphoid tissues are not equal in terms of the story they're telling," von Andrian says.

Since then, immunologists have worked out some of the chapters of that story. In a pivotal 2004 paper in *Immunity*, Makoto Iwata of the Mitsubishi Kagaku Institute of Life Sciences in Tokyo discovered that vitamin A (retinol), which is abundant in the intestine but scarce in other tissues, plays a key instructional role in T cell homing. In test-tube experiments, they found that dendritic cells from the intestinal lymph nodes convert retinol to retinoic acid, which induces T cells to make gut-homing receptors but not skin-homing receptors. Subsequent animal experiments confirmed the importance of this conversion to T cell homing: Mice starved for vitamin A had far fewer intestinal T cells than mice that consumed enough of the vitamin.

Recently, Butcher and research scientists Hekla Sigmundsdottir and Junliang Pan and their colleagues probed for a comparable molecular mechanism in the skin. "We won-

> How a newborn T cell becomes programmed to home to a particular tissue is **"a fundamentally important problem in cellular immunology."**
> —Jeffrey Frelinger, University of North Carolina, Chapel Hill

dered if a similar vitamin or metabolite that might be restricted to the skin might imprint skin homing," Butcher says. Vitamin D, which is mass-produced by skin cells in response to sunlight, "was the obvious candidate," he adds.

Butcher's team isolated lymphatic fluid from the skin of sheep, purified dendritic cells from that fluid, and found that the immune cells convert vitamin D3, the sun-induced variant of vitamin D, into its active form. In other test-tube experiments, this activated vitamin D3 induced T cells to make a receptor that helps them follow their nose to a chemoattractant in the epidermis, the skin's outer layer, the team reported in the February issue of *Nature Immunology*. An evolutionarily related chemoattractant in the

gut lures T cells using a different receptor to that tissue, Butcher points out. These studies indicate that dendritic cells can exploit a tissue's unique biochemical fingerprint—its unique mix of metabolites—to educate T cells to patrol that tissue, Butcher says.

T cells specialized for one tissue can also be retrained to patrol another area, von Andrian, HMS immunologist Rodrigo Mora, and their colleagues reported in 2005 in the *Journal of Experimental Medicine*. They cocultured T cells for 5 days with dendritic cells from the gut, spleen, or skin, which imprinted T cells for those tissues. They then washed each group of T cells and cultured them with dendritic cells from a different tissue. After 5 more days with their new instructors, "the T cell phenotype would always match the flavor of the dendritic cells they had seen last," von Andrian says. That ability to reassign T cells to new tissues may give the immune system an important degree of flexibility in fighting infection. If the pathogen stays put, the immune response is concentrated in that tissue, von Andrian says. "But if the pathogen spreads, you have not put all your eggs in one basket."

Immunologists have begun investigating whether the T cell's instructors—the dendritic cells—themselves specialize to function in a particular tissue, or whether they simply sense their environment and respond. A definitive answer is not yet in. Butcher's team found data suggesting that dendritic cells have two vitamin D–activating enzymes no matter what tissue they're from, but only in the skin do they have access to the sunlight-produced vitamin. Agace's team, in contrast, has found evidence that at least some dendritic cells are more specialized. In a 2005 study in the *Journal of Experimental Medicine*, his Swedish team reported evidence of two types of gut dendritic cells: one that has visited the intestinal wall and can train T cells to migrate to the gut, and another, of unknown origins, that can't.

Steering cells right

The new work on tissue homing is raising immunologists' hopes of specifically boosting or suppressing immunity in selected tissues. Most autoimmune diseases involve an overactive, self-destructive immune response toward a particular tissue: the pancreas in type 1 diabetes, the central nervous system in multiple sclerosis (MS), the joint in rheumatoid arthritis. Typically, treatments for such diseases dampen the entire immune system and increase the risk of infection. Similarly, stimulating the immune
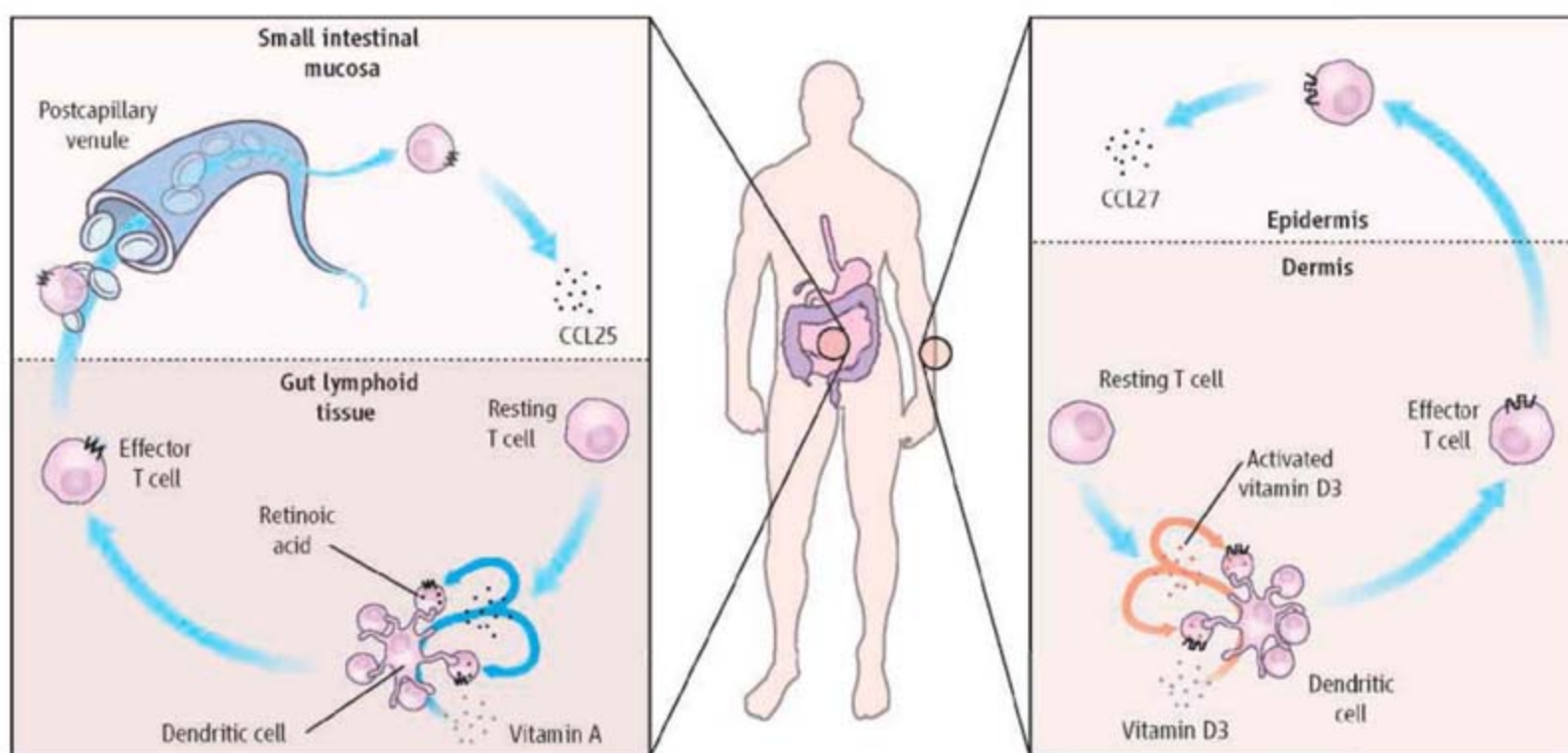
Back to the front. Dendritic cells use a tissue's characteristic metabolite—dietary vitamin A in the gut or sunlight-induced vitamin D in the skin—to educate T cells to follow their nose back to that tissue.

system nonspecifically to fight a tissue-specific tumor can increase the risk for autoimmune side effects.

That's where the new knowledge of T cell homing can help, Butcher says. Drugs that alter homing are not themselves new; in 1997, Butcher and HMS biochemist Timothy Springer co-founded a biotech company called LeukoSite, which was later bought by Millennium Pharmaceuticals, to develop drugs that block the Velcro-like interactions and molecular sniffing that help T cells find their way into tissues. Many drug and biotech companies are still pursuing that approach, which has produced a U.S. Food and Drug Administration–approved drug for MS and drugs for ulcerative colitis and Crohn's disease that are currently in clinical trials. But blocking a single receptor often fails to prevent T cell entry into tissues because the receptors involved in homing can often fill in for one another.

Drugs that alter T cell imprinting "might be a way around the problem of redundancy," Butcher says. Both gut-homing and skin-homing T cells interpret their respective signals, retinoic acid and activated vitamin D, using members of a large family of receptors that sense hormones and metabolites and directly control gene expression. Drugs that stimulate or alter these nuclear-hormone receptors already exist, and some are being tested for autoimmune diseases such as rheumatoid arthritis or psoriasis. That gives

researchers a head start, as those drugs might alter the instructions that tell T cells where to migrate, explains Butcher. "The exciting thing about imprinting is that we're just learning about its potential," he says.

The recent advances in T cell imprinting also create several possible new ways to fight disease, Agace says. Most pathogens enter the body through the surface, or mucosa, of a particular tissue, which means that a drug that directs T cells to the mucosa could enhance the cellular immune response, making vaccines more effective in warding off intruders. Other compounds could help battle localized tumors. For example, coinjecting lab-grown dendritic cells, which are already used as an antitumor therapy, with compounds modeled on retinoic acid could potentially program T cells to migrate to a gut tumor and boost the treatment's effectiveness, Agace says.

Retraining T cells could backfire by working too well, caution some immunologists. In a recent clinical trial, the MS drug Tysabri stopped abnormal T cell homing to the brain and eased MS symptoms. But it also suppressed the brain's immune surveillance system so much that a normally benign virus began reproducing in three patients, ultimately killing them.

What's more, T cells may not take instruction in all tissues, says pulmonary physician Jeffrey Curtis of the University of Michigan, Ann Arbor. Immunologists still

debate whether specific squads of T cells are assigned to patrol tissues other than the skin and gut. Researchers have been unable to find a combination of adhesion molecules or chemoattractants that lures specific T cells into lungs, he notes. But physiologist Klaus Ley of the University of Virginia, Charlottesville, who studies T cell migration in lung and blood vessel disease, disagrees: "If I project into the future, we will see more homing specificity—for gut and lung and I hope for [atherosclerotic] blood vessels."

The research on T cell homing has also now begun to merge with another hot topic in immunology: regulatory T cells, a much-touted cell type that naturally suppresses autoimmune reactions. Several years ago, Alf Hamann of Charité University of Medicine in Berlin and his colleagues reported that regulatory T cells isolated from different tissues have homing receptors like those that effector T cells sport. Now, in a March online paper in the *European Journal of Immunology*, they report that these cells, like effector T cells, can be programmed by dendritic cells, an interleukin, and retinoic acid to home to skin or gut. In theory, subpopulations of regulatory cells could therefore be prepared to target a tissue and suppress an autoimmune response. "If you could make a regulatory T cell in vitro and make it go where you want it to go, that's a cool thing," Butcher says.

–DAN FERBER
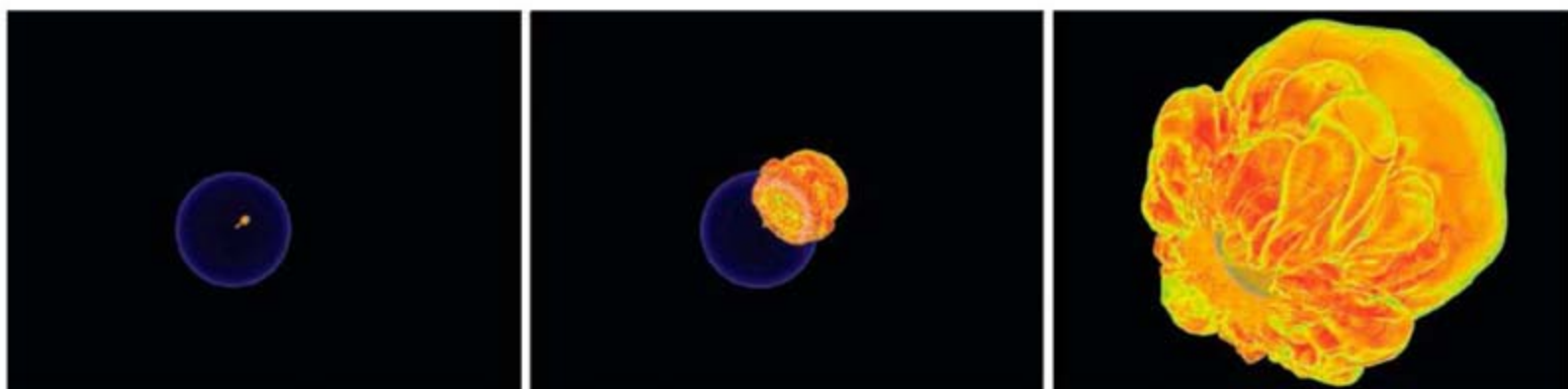
ASTRONOMY

Surveys of Exploding Stars Show One Size Does Not Fit All

Type Ia supernovae are regular enough that astronomers can use them to measure the universe. But some of the "standard candles" are breaking the theoretical mold

SANTA BARBARA, CALIFORNIA—When astronomers wish upon a star, they wish they knew more about how stars explode. In particular, experts on the stellar explosions known as supernovae wonder whether textbook accounts tell the true story—especially for a popular probe of the universe's history, the supernovae designated as type Ia.

In fact, new observational surveys suggest that cosmic evidence based on type Ia supernovae rests on a less-than-secure theoretical foundation. "We put the theory in the textbooks because it sounds right. But we don't really know it's right, and I think people are beginning to worry," says Robert Kirshner, a supernova researcher at the Harvard-Smithsonian Center for Astrophysics (CfA) in Cambridge, Massachusetts. "We keep saying the same thing, but the evidence for it doesn't get better, and that's a bad sign."

Kirshner was among more than 100 experts on stars and their explosions who gathered to discuss their worries last month at the Kavli Institute for Theoretical Physics at the University of California, Santa Barbara.* General agreement emerged that the textbook story "is a little bit of 'the emperor has no clothes,'" as Lars Bildsten, an astrophysicist at the Kavli Institute, put it. "There's a lot of holes in the story."

Understanding type Ia supernovae has become an urgent issue in cosmology, as they provide the most compelling evidence that the universe is expanding at an accelerating rate. That acceleration, most cosmologists

conclude, implies the existence of a cosmic fluid called "dark energy" that exerts a repulsive force countering gravity.

In the textbook story, type Ia explosions occur in binary systems where a worn-out star known as a white dwarf siphons matter from a nearby companion. When the planet-sized dwarf accumulates enough mass to exceed the Chandrasekhar limit—about 1.4 times the mass of the sun—its density becomes great enough to ignite thermonuclear fusion, blowing itself to smithereens.

Because all white dwarfs presumably blow up the same amount of mass, they should all be equally bright at any given distance, and so their apparent brightness should diminish with distance in a predictable way. Faraway type Ia supernovae are dimmer than expected, however, suggesting that the universe's expansion rate has been speeding up.

But figuring out exactly what dark energy is will require a precise gauge of its effect on the expansion history of the universe. And type Ia supernovae are not yet well enough understood for analysis of their brightness to provide the needed precision, experts say. "We do not know the details," says Alex Filippenko of the University of California, Berkeley. "There is still a lot of controversy about what exactly is going on in a Ia."

Several speakers during the Santa Barbara conference noted problems with the textbook view. For one, astronomers have long realized that not all type Ia's explode with the same brightness. Instead, the brightest are several times as luminous as the dimmest. Type Ia explosions in old, elliptical

galaxies appear dimmer, on average, than explosions in younger galaxies. It may be that such differences reflect different pathways leading to explosion, hinting that type Ia supernovae come in two distinct flavors. "There is now very strong evidence that … there are very likely two populations of type Ia supernovae," said Bildsten.

Corrections for brightness differences can be made based on the color of the explosion's light and how rapidly it dims. Such fixes were good enough to establish accelerating expansion but not for pinning down dark energy's properties precisely. That will require answers to several nagging questions, including the nature of the white dwarf's companion and the mechanism of the explosion.

The good news from the conference is that several computer simulations seem to show that a 1.4-solar-mass white dwarf can indeed explode like a bomb, although various models differ in their details. In some models, a wave of fusion burns slowly through the star (a process known as deflagration), ultimately detonating the fast-burning explosion that mimics a hydrogen bomb. In the star, however, the elements fused are carbon and oxygen, the elements believed to make up the bulk of the white dwarf type Ia progenitors.

Immediate detonation of the entire star in a rapid shock-wave blast is unlikely because it would convert nearly all the material into an isotope of nickel (which eventually decays to form iron). Because intermediate-weight elements (such as silicon) are found in type Ia debris, some of the burning must be slower.

A deflagration model discussed at the conference by Wolfgang Hillebrandt of the Max Planck Institute for Astrophysics in Garching, Germany, seems able to produce an explosion, but only if deflagration begins at multiple points within the star. Another approach, presented by Don Lamb of the University of Chicago in Illinois, showed how a bubble of fusion beginning inside the star can burst out through its surface and then, confined by the star's gravity, wrap

Kaboom! Computer models show ways stars might explode but not what primes them for the blast.

around the star in all directions, until encountering itself on the other side (see figure, p. 194). When the fusing material collides with itself, a jet of material fires back down into the star, detonating the full fusion explosion, a new three-dimensional computer simulation shows, confirming the basic picture seen in earlier two-dimensional models.

But, as Kirshner pointed out, simulating an explosion is one thing. It remains to be seen whether the models can replicate the energy and mix of elements actually seen in various type Ia explosions. And these models assume that a 1.4-solar-mass white dwarf is conveniently available and poised to explode, yet nobody knows exactly how white dwarfs reach that point, or whether there are enough of them to account for the observed rate of explosions. In fact, most observed white dwarfs are typically only a little heavier than half the mass of the sun, far below the explosion point.

In the standard story, white dwarfs reach the mass limit by accreting hydrogen from a companion star. But the accretion must occur at a "just right" rate—too fast, and it will be blown away by smaller explosions before reaching the bomb mass.

Furthermore, if white dwarfs really explode by accreting hydrogen from a companion, leftover hydrogen should be visible in the supernova remnant. But sensitive observational searches have failed to find the hydrogen. "I think this lack of hydrogen is a very, very serious issue," said Filippenko.

The missing hydrogen leads some experts to speculate that the companion star is not an ordinary hydrogen-rich star but something else—perhaps even another white dwarf. But searches find few double-dwarf systems likely to become supernovae. The Supernova Ia Progenitor Survey at the European Southern Observatory in Chile has observed more than 1000 white dwarfs and has found only two double-dwarf systems, Ralf Napiwotzki of the University of Hertfordshire, U.K., said at the conference.

In one, the total mass of both dwarfs didn't reach the explosion threshold, and they wouldn't merge for 25 billion years, anyway. The other double dwarf falls just

short of bomb mass. "At the moment, we can't say we have a clear-cut supernova Ia progenitor," Napiwotzki said. But deeper searches may find more candidates, he added.

If double dwarfs do merge and explode, their combined mass could exceed the Chandrasekhar limit, producing an unusually bright explosion. And in fact, one such unusual explosion was spotted in 2003 and reported in *Nature* last year by the Supernova Legacy Survey, an international project using the Canada-France-Hawaii telescope on Mauna Kea.

Supernova 2003fg looks like a type Ia, said Andrew Howell of the University of Toronto, Canada, but glows with more than double the median Ia brightness. Its brightness and energy output suggest a combined mass of more than two solar masses, implying (among other possibili-

What next? Uncertainties in supernova surveys could muddle efforts to determine the nature of dark energy—and thus the fate of the universe.

ties) a double-dwarf explosion or the growth of a single white dwarf to larger than the expected maximum mass. Many experts find it hard to envision a single dwarf growing that fat, but neither has current theory established that the merger of two dwarfs would produce the observed features of a type Ia explosion.

In any case, freak explosions such as 2003fg are just the sort that could contaminate supernova data needed to determine whether dark energy is the residual energy of empty space incorporated by Einstein into his theory of relativity as a "cosmological constant." If it is, the ratio of the dark energy's pressure to its density would be exactly −1, at all times and places throughout the universe. (That ratio, known as the equation of state, is negative because the pressure is negative, confer-

ring the dark energy's repulsive effect.)

If the ratio is greater than −1, dark energy could be a new sort of field, sometimes called quintessence, that changes its strength over time. A ratio less than −1 suggests an entirely weird "phantom" energy that would someday rip the universe to shreds (See figure below and *Science*, 20 June 2003, p. 1896).

Current efforts to gauge the equation of state using supernovae are all consistent with −1 but not sensitive enough to detect small deviations. At the conference, Mark Sullivan of the University of Toronto reported a Supernova Legacy Survey analysis of 250 supernovae giving a value of −1.02, but with an error range including −1. Michael Wood-Vasey of CfA, presenting for another supernova survey known as ESSENCE, reported −1.05, based on more than 170 supernovae, but again with uncertainties large enough to include −1.

Reducing such uncertainties further is a prime goal of several supernova-search satellite missions to probe dark energy that will be competing for funding, as described in last year's Dark Energy Task Force report prepared for NASA, the National Science Foundation, and the Department of Energy (www.science.doe.gov/hep/DETF-FinalRptJune30,2006.pdf). But some experts doubt that supernova theory will ever be good enough to identify small deviations from −1, even with thousands of supernovae observed from a dark-energy satellite. (Some of the proposed missions, however, would measure both supernovae and other features, such as gravitational-lensing effects, that could help narrow the uncertainties.)

In any event, better supernova data could still be useful to cosmologists, Bildsten pointed out. "If there's really two populations, you might decide that one of those populations isn't so good, and if it's in this type of galaxy or that, you don't use it for your cosmology," he said. "Maybe that's helpful information."

But whatever help supernovae can provide will still depend on plugging the worrisome gaps in current textbooks accounts, Kirshner said, and answers to many critical questions remain elusive. "I wouldn't say it's a crisis," he said. "But if you ask, 'Are the pieces falling into place?' I'd say the answer is no."

–**TOM SIEGFRIED**

Tom Siegfried is a writer in Los Angeles, California.

In bloom. K. B. Saxena (*right*) and colleagues bred countless varieties of pigeon pea to create new hybrids.

AGRICULTURE

The Plant Breeder and the Pea

K. B. Saxena has spent his career trying to boost yields of pigeon pea, a crop relied on by hundreds of millions of marginal farmers. At last, he's succeeded

When he decided on his life's work as a plant breeder, K. B. Saxena made an unlikely choice. The year was 1974, and new varieties of rice and wheat were boosting production and cutting hunger around the world. With a newly minted Ph.D. from one of India's top agricultural universities, Saxena could have worked on any of these blockbuster crops. Instead, he picked a gangly, unrefined plant called pigeon pea.

Although still barely known in the West, pigeon pea (*Cajanus cajan*) is the main source of protein for more than a billion people in the developing world and a cash crop for countless poor farmers in India, eastern Africa, and the Caribbean. This hardy, deep-rooted plant doesn't require irrigation or nitrogen fertilizer, and it grows well in many kinds of soil. "It's such an important crop, and it had been neglected," Saxena says.

During a 30-year career at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in Patancheru, India, Saxena helped create nearly a dozen kinds of pigeon pea that mature sooner and resist diseases better than do traditional varieties. Yet the big prize—high-yielding hybrids—never seemed within reach. "People had lost hope that yield could improve," says Saxena, who narrowly escaped being laid off a decade ago and barely managed to keep his program going during hard times at ICRISAT.

Now, hope is back. Two years ago, Saxena's group finally succeeded in creating the first commercially viable system in the world for producing hybrid legume seed. It couldn't have come at a better time: India faces a pigeon pea shortage severe enough that the government banned exports of it and other so-called pulses last year. Last month, ICRISAT announced that one of its most promising hybrids can achieve yields nearly 50% higher than those of a popular variety. "This will become the forerunner of a pulses revolution in India," predicts M. S. Swaminathan, a plant breeder considered one of the chief architects of the original green revolution. The first seeds should reach farmers next year, and Swaminathan himself is working on a project to make sure even the poorest can afford them.

Deep roots

Saxena was inspired to become a plant breeder when he was in high school. His older brother, a maize breeder, would take him into the research fields and explain what he was doing. "All that stimulation came from my brother," Saxena says. "He encouraged me a lot." And with the green revolution at its height, plant breeding was a hot field. After finishing a Ph.D. in cereal grains, Saxena joined ICRISAT in 1974, which had been founded just 2 years earlier to improve five semiarid tropical crops: sorghum, pearl millet, chickpea, groundnut, and pigeon pea.

There wasn't much competition to work on pigeon peas, Saxena recalls. Crops took 6 to 9 months to mature, slowing the pace of research. And they grew to 2 to 3 meters tall, their pods covered in a sticky gum. "It will spoil all your clothes in an hour," Saxena says. "No one wanted to work on such a dirty crop."

But sensing an opportunity—and loving the dahl made from pigeon peas—he plunged in. By the 1980s, the small team of plant breeders at ICRISAT—together with researchers at the Indian Council of Agricultural Research (ICAR)—had developed early-maturing varieties that can be harvested in only 3 months. That meant an entire crop of nitrogen-fixing pigeon pea can be planted before the wheat crop in northern India, helping to restore fertility to the soil. New varieties also featured improved resistance to fusarium wilt and the dreaded sterility mosaic virus known as "the green plague." But yields hardly budged, rising to an average of 700 kilograms per hectare.

The way to smash through the yield barrier is by creating plants with hybrid vigor. This is a well-known phenomenon in which the first generation of offspring exhibit vastly superior traits—yield, or overall health, for example—than those of either parent. The process starts with picking the best plants from each generation and breeding them so that all the progeny of each have dependable traits, then crossing them. This is relatively straightforward and can be done by hand in the greenhouse.

But making enough hybrid seed to sell requires an easy way to prevent plants of each parent variety from fertilizing themselves. (Each plant carries both male and female sex organs.) Breeders like to create so-called male sterile plants that can't make viable pollen but can still be fertilized by pollen from certain other varieties. In corn and rice, varieties had been bred to produce sterile pollen by the 1980s.

Breeding sterile plants in pigeon pea and other legumes has proven much more difficult. For starters, the male and female parts exist within the same flower. That means researchers must pollinate the delicate ovaries by hand, and sometimes only a few percent can be successfully fertilized. This and other challenges kept hybrids off the agenda of most legume breeders. "It's theoretically possible,

but it's hard to do," says Harbans Bhardwaj, a plant breeder at Virginia State University, Petersburg, who has worked with pigeon pea.

Proof of principle came in the late 1980s. Working with ICAR institutions, Saxena and his ICRISAT colleagues found varieties with nuclear genes that conferred male sterility. Hybrids from this line boosted yield by up to 30% and did well in field trials. But the males weren't always sterile—sometimes just a fraction of the male flowers lacked pollen. Because not all the seeds produced from them were high-yielding hybrids, companies were not interested in commercializing the plants.

Other setbacks put the project in jeopardy, too. In the late 1990s, ICRISAT began to have major budget problems (*Science*, 2 January 1998, p. 26), and management decided to drop pigeon pea research. Saxena pleaded his case, but the pigeon pea team was still cut deeply; the three other breeders were laid off, as were seven of 10 technicians. After drumming up external funding from companies, Saxena rebuilt the program.

The big advance came from finding plants with genes in their cytoplasm that confer male sterility. Unlike nuclear genes, which are segregated during meiosis, the cytoplasm is passed down through the eggs to all the progeny, so the offspring of plants with the particular cytoplasmic genes will all be male-sterile.

Some hints of cytoplasmic male sterility had come early on, when breeders managed to produce plants with sterile pollen by fertilizing wild pigeon peas with pollen from a cultivar. But the plants also had sterile ovaries. Saxena expanded this effort with other wild species. Each year, the group made thousands of experimental crosses and planted the seeds. Every plant was inspected when it flowered—an onerous task, as a hectare can contain up to 60,000 individual plants. In March 1996, a research assistant struck gold when he located plants with no pollen in any of their flowers—a sign of cytoplasmic male sterility. He dashed off on his motorbike to relay the news. "It was very exciting," says Saxena. "People were really smiling in the field." Further testing revealed that the sterility was indeed passed on through the cytoplasm.

Still, success came slowly. Some lines created plants that were male-sterile when grown in winter but would somehow produce fertile pollen in the summer. "Sometimes we got frustrated," Saxena says. Out of five lines

they developed, only one turned out to be stable enough. In order to perpetuate these plants, they also had to find a maintainer line—a nearly identical plant with pollen that would fertilize the plants but yield seed that would grow up into male-sterile plants. The final component was a restorer line, a variety that can fertilize the male-sterile plants so that the progeny will bear seeds. New traits, such as disease resistance, are bred into either or both of the parent lines to produce hybrid seed for farmers.

By 2004, the system was up and running (*Euphytica*, October 2005, p. 289). "This is a breakthrough in plant breeding," notes

Seeing green. Extra yield from hybrid varieties of pigeon pea could raise the income of farmers, such as these pea farmers in Eastern Kenya.

Latha Nagarajan, an agricultural economist at the International Food Policy Research Institute in Washington, D.C. "The possibilities are endless."

Starter seed

Of some 300 hybrids tested so far, the best—called ICPH 2671—yields up to 3 tons of pigeon peas per hectare—48% more than a widely used variety known as Murati. "It's a quantum jump in yield," says Swaminathan, who thinks that yields could even be doubled with improved cultivation and pest management. He notes that legumes require careful attention to phosphorus, and Indian soil is often poor in micronutrients that pigeon pea needs, such as zinc and boron, so

educating farmers about soil nutrition will be important.

In addition to yield, Saxena says that this hybrid also does better against drought and resists diseases better than do the standard lines. "This system that Saxena has developed will benefit the small subsistence farmers and consumers," says Sharad Phatak, a horticulturist at the University of Georgia, Tifton. If 15% to 20% of the acreage is planted with hybrids, he reckons, it might take care of the pigeon pea shortage.

There is, however, a downside: Unlike traditional varieties, hybrid seed must be bought every year, because only the first generation has the hybrid vigor. Most of those seeds will come from companies, which makes some observers worry that small farmers won't be able to afford them. ICRISAT has provided the male sterile system, which is in the public domain, to a consortium of 15 Indian seed companies so that they can create their own hybrid pigeon pea varieties. Several companies are also preparing to sell ICPH 2671, and Saxena estimates that the seed will cost about $3.25 a kilo, about 50% more than standard cultivars. He says it's likely that some government agencies will sell the seed at half price to poor farmers.

Swaminathan isn't taking any chances, however. His foundation, based in Chennai, is beginning a project to train women to produce the hybrid seed themselves from ICRISAT seeds. Beginning in June, agronomists will go to villages about 180 km south of Chennai and teach some 100 women, mostly the wives of subsistence farmers. The goal is for them eventually to sell hybrid seed in their neighborhoods. "The principle is social inclusion and technology access for all," Swaminathan says. "You can keep the cost of seed low and increase employment in villages."

Meanwhile, Saxena, now 58 and 2 years from retirement, spends half his time promoting the sterile lines to breeders at universities and companies, as well as encouraging farmers to try them out. He's also involved in promoting the use of pigeon pea in other countries, such as China, where it's used in several ways but mainly as a quick fix for soil erosion. Although Saxena's work may not trigger the dramatic agricultural revolution that he witnessed at the beginning of his career, it could still improve the lives of hundreds of millions of people.

–ERIK STOKSTAD

COMMENTARY

Dance through time
203

Monitoring the environment in India
204

LETTERS | BOOKS | POLICY FORUM | EDUCATION FORUM | PERSPECTIVES

LETTERS

edited by Etta Kavanagh

NIH Funding: What Does the Future Look Like?

STRATEGIES TO SUPPORT AND FUND THE NATIONAL INSTITUTES OF HEALTH (NIH) NEED TO BE AIMED not only at mitigating the long-term effects of the current budget woes, as Elias Zerhouni eloquently articulated in his Policy Forum "NIH in the post-doubling era: realities and strategies" (17 Nov. 2006, p. 1088), but also at creating a sustainable model for what NIH does best: individual investigator-initiated research. Although the numbers cited by Zerhouni regarding unsolicited, independent grant mechanisms are reassuring, the Federation of American Societies for Experimental Biology (FASEB) is concerned that slow erosion of this pool, by internal and external pressures, may ultimately diminish NIH's ability to fund this extraordinarily successful pathway to discovery.

Congress has been supportive of the establishment of new centers, as Zerhouni himself has said, as well as initiatives like the growth of the Common Fund. The fiscal year 2007 funding resolution specifically directs funds for the National Children's Health Initiative, a dangerous precedent that could leave some NIH Institutes struggling to fund additional research. Regardless of the merits of any of these projects or mechanisms, it is indisputable that over time the piece of the NIH portfolio dedicated to independent investigator-initiated research will grow smaller. We need to very carefully consider the wisdom of such allocations at a time of extreme fiscal constraint.

FASEB fully agrees with Zerhouni's analysis pointing to increases in application numbers, flat federal budgets, and inflationary pressures as the major factors in creating the current anxiety-inducing environment. Moreover, we applaud and echo many of the strategies proposed: protecting new investigators, improving peer review, and communicating the value of research to Congress and the public are basic principles the scientific community should give the utmost consideration. However, as NIH contemplates how to spend what little funds they have available and as Institutes go through the "strategic priority setting" described by the Director, it is imperative that they contemplate the long-term effects on the ability of creative scientists to pursue great ideas. Individual investigators have sustained and driven our scientific enterprise for centuries, and the scientific enterprise, in turn, needs to sustain their productivity.

LEO T. FURCHT

President, Federation of American Societies for Experimental Biology (FASEB), 9650 Rockville Pike, Bethesda, MD 20814–3998, USA.

IN HIS POLICY FORUM, ELIAS ZERHOUNI attributes the current grant crisis to the inflation in research project grant values (≈40% since 1998), increased demand resulting from an expansion of the faculty ranks nationwide, and a flat NIH budget. For many, these facts are obvious, but so too is the fact that the NIH budget did double and the enhanced grant values were administratively permitted by the very agency that is now struggling to deal with the unanticipated consequences of the decision. One wonders, therefore, whether sufficient long-term planning accompanied the euphoria resulting from the increase in funding for biomedical research that resulted from the doubling period (1998–2003).

The most commonly expressed concern by academic faculty is the administrative status of the R01-equivalent grant (*1*). Although Zerhouni restated the agency's commitment to investigator-initiated research, the data posted by the Office of Extramural Research paint a different picture (*2*). The funding, in dollars, for R01-equivalents increased between 1998 and 2005, but the percentage these dollars represent relative to the total NIH grant budget has been steadily declining. As of the end of fiscal year 2005, the R01-equivalent represented 12% less of the overall grant budget relative to 1998, the year the NIH budget began to double. During this same period, there has been an increase in the total number of awarded grants, but these increases did not include a sustained increase in the total number of R01-equivalents awarded. There was a brief period of increase in the awarding of R01-equivalents between 2000 and 2003, but this increase was all but erased by 2005. This means that most of the sustained increases in the number of NIH awards were in non–R01-equivalent categories. Moreover, the number of new investigator R01-equivalent grants has remained essentially unchanged since 1981, including the doubling period 1998–2003. Thus, although the doubling of the budget significantly increased the total funds for research, the increases appear to have been targeted away from new investigators and, to a lesser extent, away from the R01-equivalent itself. Collectively, these observations raise questions about the NIH's commitment to investigator-initiated research, i.e., the R01-equivalent. The problem today is not that there are too few funds available—we have nearly twice the funds we had circa 1998. Rather, we have to recognize what grant funding means to the careers of academic faculty and that the current crisis is destroying the careers of people and not just canceling grants, and we must really rethink how best to distribute the funds we do have to sustain the careers of institutional scientists nationwide. We must, in short, manage the money better and distribute funds differently as we look forward. An increase in the NIH budget should have led to a sustained expansion of research conducted by individual investigators. Instead, many of us are tearing down significant elements of our programs just to survive.

MILTON H. WERNER

CREDIT: PETER HOEY
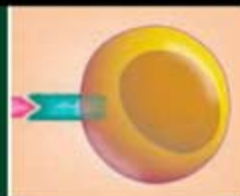
Laboratory of Molecular Biophysics, The Rockefeller University, New York, NY 10021, USA. E-mail: mwerner@rockefeller.edu

References and Notes

1. Defined by NIH as the R01, R23, R29, and R37 mechanisms; see http://grants1.nih.gov/grants/award/award.htm.
2. See http://grants1.nih.gov/grants/award/award.htm.

IN HIS POLICY FORUM, ELIAS ZERHOUNI rejected the idea that the NIH Roadmap detracts resources from unsolicited investigator-initiated proposals (R01), arguing that this program is only a small percentage of the total NIH budget and that it supports some R01s grants. This is not the case. The type of R01s to which Zerhouni refers responds to various special initiatives and program announcements (PAs) covered under the umbrella of the Roadmap. Since the topic of investigation for these R01s is decided by the NIH administration, they do not fit the definition of unsolicited investigator-initiated proposals. On the contrary, the success rate of true investigator-initiated grants has decreased dramatically, while NIH continues to promote multiple special programs through the R01 award mechanism. Consequently, if an appropriate balance between NIH-solicited- and investigator-initiated proposals is lost, the scientific community risks disenfranchisement of its role in guiding scientific directions.

From fiscal years (FY) 2003 to 2005, the number of dollars awarded to R01s has decreased by $142 million for unamended applications and by $131 million for renewal applications (1). In stark contrast, the budget for the Roadmap has doubled by $203 million from FY 2002 to FY 2007 (2). The biodefense program is also an expanding priority in the NIH portfolio, with a total budget increase of $110 million for FY2007 and $182.6 million spent on construction inside the NIH campus. Furthermore, NIH promotes, at the cost of several hundred million dollars, a number of special funding initiatives that favor "large-scale fishing expeditions," such as genomic and proteomic analysis. As an example, the Cancer Genome Atlas program has a budget of more than $100 million. Therefore, in spite of a flat budget, NIH still manages to increase resources for multiple programs. Disappoint-

ingly, President Bush's FY2008 budget request for NIH represents an actual cut of 1.7% (3), forecasting a further decline in success rates for R01s. Furthermore, payline levels have remained significantly lower than they were in the past. This is creating disturbing discrepancies in the grant review process. Unlike a few years ago, today, applications ranked as scientifically outstanding may fall in a percentile range that is not funded. Applicants and reviewers no longer see a connection between scientific merit and actual awards, and meritorious projects suffer fatal delays in funding. This is deleterious to the scientific progress. Thus, effective adjustments should be made to improve payline levels and to truly address the problem of the loss of funding of unsolicited R01 proposals.

Zerhouni outlines multimillion-dollar funding strategies to preserve future generations of young investigators, which also utilize the R01 mechanism. The problem here is that new and existing R01s are currently slashed with severe administrative cuts, endorsed by NIH to fund more grants. The new investigator typically has one R01, upon which the salary of personnel, his/her stipend, and supplies all depend. In only four years, however, the "no longer new" investigator will compete with senior scientists for renewal. Chances of success are not very high. This scenario applies to senior investigators as well, who see severe reductions of their budget. Thus, since there is no strategy to keep young investigators in business once they mature, the initiatives in favor of this category of scientists are expensive, but short-sighted. The anemic budgets awarded by NIH hamper everyone's ability to realistically perform research. Hence, the vicious cycle: less money is awarded per single grant, and fewer grants are available per single investigator, which begs for more grant writing efforts. This further saturates the system by increasing the load of submissions and negatively impacting upon success rates. Zerhouni does not explain how his proposals solve this ill-fated Catch-22.

Zerhouni states that the mission of the special funding initiatives promoted by NIH today is to preserve high-risk, high-impact science. I believe that this kind of science is best served when it comes from within the scientific community. In the postdoubling era, the expansion

of the Roadmap and of similar funding initiatives comes at the price of curtailing vital resources for the unpredictable, creative thinking of scientists of all categories. Thus, the question is, who should decide what area of investigation is worth pursuing and which is destined to advance science? Unfortunately, the NIH leadership falls short in the credible intention of maintaining an acceptable balance between its own special funding initiatives and investigator-initiated proposals. Changes could and should be made. For example, call for an open discussion that will allow the active scientific community to have a more direct role in the way NIH prioritizes the distribution of its budget and reduce as much as necessary the number of special funding initiatives. These are the kind of measures that should help put some money back where it truly belongs: in the pockets of free-thinking scientists at all types of academic institutions.

MARIA LAURA AVANTAGGIATI*

Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC 2007, USA, and Permanent Member of the Drug Discovery and Molecular Pharmacology Study Section, DMP, NIH. E-mail: ma364@georgetown.edu

*The opinions contained herein represent the opinions of the individual author and do not necessarily represent the opinions, policies, or practices of Georgetown University or the Lombardi Comprehensive Cancer Center.

References and Notes

1. G. H. Mandel, *Science* **313**, 1387 (2006).
2. See http://officeofbudget.od.nih.gov/pdf/Press%20info%20final.pdf.
3. E. Marris, *Nature* **445**, 572 (2007).

Response

I WOULD LIKE TO THANK THOSE WHO TOOK time to respond to my Policy Forum about the challenges facing NIH and its scientists. Some would prefer to dedicate the NIH budget to unsolicited investigator-initiated R01 research, and others feel that NIH should increase its focus on larger programs, including clinical trials and translational research. Clearly, the answer lies somewhere in between. If NIH strays too far from fundamental research or fails to fund more applied science, we will lose what makes NIH successful—new scientific insights and knowledge that improve health. NIH thrives by promoting diverse approaches to science and health, based on wide-ranging input and review by scientists. However, the majority of our investments are and should remain focused on basic research, i.e., long-term research that has a high potential for impact.

NIH leadership is deeply concerned about R01 funding levels and dropping success rates. Of note, the difference in success rate between solicited and unsolicited grants has remained

constant through the doubling, however, unsolicited grants have always been higher. We must preserve the strength of the unsolicited investigator-initiated R01 program as the backbone of discovery. Today, the R01 program is the largest program at NIH, in both dollar amounts and number of grants funded. In absolute dollars, the R01 program has grown the most—the total number of dollars awarded to R01s has almost doubled in recent years, growing from $5.3 billion in 1998 to $10.1 billion in 2006. The rate of increase slowed following the doubling of the budget, yet during the period between 2003 and 2005 the R01 total funding still increased by $500 million.

However, for NIH's overall mission to succeed, NIH complements R01s with other investments to fill knowledge gaps and fund emerging areas of science. We consistently reach out to the scientific community to help inform our decisions as to where the R01 mechanism cannot meet the needs of science or scientists. The NIH Roadmap stimulated much discussion in this regard. It is not one large initiative created from the top down, but is composed of several hundred highly competitive awards that emerged from extensive discussions with working scientists at large.

We also recognize that it is important that the size of grant awards rise in parallel with biomedical research inflation. This long-term strategy allows researchers to maintain their purchasing power. This said, in the face of unrelenting budget constraints, we made the difficult decision to reduce all noncompeting Research Project Grant (RPG) commitments by 2.35% in FY 2006 and 3.0% in FY 2007, freeing up an estimated $1.35 billion over the lifetime of these grants to invest in more competing RPGs, including R01s.

Science is about people. And current constraints are putting some categories of scientists at particular risk. We strongly advocated for the next generation of scientists through new grant programs for new investigators (1), and this year we are working to support first-time and more established investigators who are at risk of losing their laboratories because they receive review scores on their R01 renewal applications

that are near but beyond the payline (2).

We are delighted that Congress increased the FY 2007 budget by $687 million. In total, we are now able to support more than 10,000 competing RPGs, 1000 more than in FY 2006. Yet, the nation still faces daunting scientific and health challenges. We should strengthen our common efforts to speak out about the importance of medical research to the nation's health. The fundamental problem for the research community is the loss in purchasing power relative to the ever-increasing scope of research needs and demands. As scientists, we need to express a broad, compelling vision for the future of science, health, and medicine if we are to sustain NIH's noble mission. A recent vote of renewed confidence in NIH by Congress through the NIH Reform Act of 2006 permits some optimism.

ELIAS A. ZERHOUNI

Director, National Institutes of Health, 1 Center Drive, Bethesda, MD 20892–0148, USA.

References

1. See http://grants.nih.gov/grants/new_investigators/index.htm.
2. See http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-047.html.

Are There Too Many Scientists?

IN THEIR LETTER "DECLINES IN FUNDING OF NIH R01 research grants" (8 Sept. 2006, p. 1387), H. G. Mandel and E. S. Vesell reiterated the widespread concern that the low level and recent decline in funding NIH R01 grants has long-term negative consequences for the future quality of science. Weinberg has also commented that the shift away from funding small, independent research groups is detrimental to the progress of science (1). I agree with these sentiments, but there is another side to the funding issue: Are there too many people trying to do science?

In both the United States and Canada, the budgets of the major national agencies that fund biomedical research have more than doubled during the past decade. At least some of that money has made it to the pool that supports investigator-initiated research, yet excellent proposals from excellent researchers are still not being funded, and the situation has become worse in the past year (2). We could lobby for another doubling of the funding, but I see no reason to expect the outcome to be different: The total amount of science would increase, but the probability of continued funding and the perceived desirability of science as a career would not.

Is there a point at which a society is doing enough science? Is science so important that we should always want to increase the rate at

which we do it? Or is it like almost every other government-funded activity, where the proponents always want more even if we are not sure that more is better? If there were only 100 scientists applying for R01 grants each year, there would be widespread agreement that this is not enough to sustain a vibrant research enterprise; if there were a hundred million, even the most ardent supporters of research would agree that this is too many. The right number, or range of numbers, must be somewhere in between. We could lobby to keep doubling the funding and hope to reach the point where all of the good science was being funded. However, if funding for science is like funding for medical care, education, or war, there is no precedent to expect that increases in funding will ever match the ability to spend the funds, for better or worse. Alternatively, or in addition, if we don't like the model of funding that we have created, we should debate the merits of limiting the demand for research grants rather than just increasing the supply of money.

RICHARD A. COLLINS

Department of Molecular and Medical Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada.

References

1. R. A. Weinberg, *Cell* **126**, 9 (2006).
2. J. Giles, M. Wadman, *Nature* **443**, 894 (2006).

Fishing for Good News

IN THEIR REPORT "BIOMASS, SIZE, AND TROPHIC status of top predators in the Pacific Ocean" (15 Dec. 2006, p. 1773), J. Sibert *et al.* consider the effects of fishing on certain top predators in the Pacific Ocean. The bad news in their Report is that two large and extremely valuable tuna species have declined steadily and that the proportion of large fish has been reduced by 80%. The good news is that two smaller and significantly less valuable tunas and a commercially irrelevant shark have either not declined or have increased slightly. Sibert *et al.* conclude that the impact of fisheries on Pacific top predators is "not catastrophic" and that fishing has had only "minor impacts" on pelagic ecosystems. Both conclusions merit reexamination.

When are we to consider the impacts of a fishery to be catastrophic? Fisheries managers from eastern North America, unfortunate witnesses to the collapse of some of their most valuable top predator fisheries, no doubt look enviously west at the Pacific. But must our definition of a fisheries catastrophe be an empty ocean? We are all starved for good news from the seas. However, I would argue that we have little reason to rejoice over news that the Pacific's most desirable tunas are becoming smaller and less numerous as their diminutive, less valuable, and less palatable

counterparts increase.

It is impossible to evaluate Sibert *et al.*'s second conclusion because no data are presented on how fishing affects the Pacific "ecosystem" beyond their top predator assessments. To understand the impact of a fishery upon an ecosystem, we need information on how it influences lower trophic level organisms, primary productivity, and ecosystem processes. This is a tall order for the Pacific Ocean. However, because it has long been known that top predator depletions can dramatically reorder marine ecosystems (*1, 2*), it is difficult, without evidence, to be convinced that removing 50 million tons of the Pacific's top predators has had only minor effects.　**DOUGLAS J. MCCAULEY**

Hopkins Marine Station, Stanford University, Pacific Grove, CA 93950, USA.

References

1. R. T. Paine, *Am. Nat.* **100**, 65 (1966).
2. J. A. Estes, J. F. Palmisano, *Science* **185**, 1058 (1974).

Response

MCCAULEY'S CONCERNS REST ON INACCUrate assumptions and reflect unfamiliarity with fisheries management and tuna fisheries in particular. Contrary to McCauley's implied claim that species other than bigeye and yellowfin tuna are of little value, skipjack tuna is the mainstay of the global canned tuna industry, providing an inexpensive source of healthy protein for millions of people (*1*). Furthermore, blue shark is an abundant and important top predator, with annual longline catches in the western and central Pacific of 40,000 to 100,000 tons over the past decade (*2*), easily the largest component of the nontuna catch.

Reexamination of the structure and function of the oceanic ecosystem of the Pacific Ocean should be a top research priority (*3*). The ecosystem effects of fishing are generally difficult to evaluate, and there are no widely accepted indicators of them (*4*). More extensive investigations than ours of ecosystem structure in the open ocean have shown that lower trophic levels are surprisingly insensitive to reductions in top predator biomass; the expected control of food web structure by top predators has been difficult to establish (*5*).

Decreased average size and reduced biomass are two textbook examples of the effects of fishing on exploited populations. Exploiting a fish population will reduce both abundance and the average sizes of individuals in the population. Further, optimal management for yield of a single species will substantially reduce the size of the exploited population. Our results indeed show that the abundance of fish larger than 175 cm is about 20% of what it might have been in the absence of fishing. However, fish of this size were never very abundant, comprising only about 4% of the biomass in the 1950s (fig. S5). The levels to which abundance of some species have been reduced are capable of supporting maximum sustainable yield (MSY), even though they are less than 40% of what the biomass might have been in the absence of fishing. Of more immediate concern is that the levels of fishing mortality to which some species are currently subjected will ultimately cause declines to well below MSY abundance. Whether MSY is an appropriate standard for an ecosystem approach to fisheries is another question that is under investigation and debate in many ecological research centers.

The purpose of our article was not to contrive good news. Rather, we strived for an accurate view of the status of tuna stocks and their management in the Pacific Ocean. Our purpose was to show that the interactions between fisheries and the pelagic ecosystem are complex and not easily summarized in flashy, often exaggerated, headlines (*6*) proclaiming the collapse of ocean ecosystems and the end of fishing. Effective conservation measures must consider this complexity. We show that some stocks of tuna are being exploited at excessive rates. This fact and potential remedies for the problem are well known to scientists serving the agencies charged with regulating tuna fisheries. These agencies need to fulfill their responsibility to implement effective science-based regulations to maintain these stocks.

JOHN SIBERT,[1] JOHN HAMPTON,[2] PIERRE KLEIBER,[3] MARK MAUNDER[4]

[1]Joint Institute for Marine and Atmospheric Research, University of Hawaii, Honolulu, HI 96822, USA. [2]Oceanic Fisheries Programme, Secretariat of the Pacific Community, BP D5, Noumea 98848, New Caledonia. [3]Pacific Islands Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2570 Dole Street, Honolulu, HI 96822, USA. [4]Inter-American Tropical Tuna Commission, 8604 La Jolla Shores Drive, La Jolla, CA 92037, USA.

References and Notes

1. P. Williams, C. Reid, WCPFC-SC2-2006/GN WP-1, 2nd Annual Session of the Western and Central Pacific Fisheries Commission Scientific Committee, Manila, Philippines, 7 to 18 August 2006 (www.wcpfc.int/sc2/pdf/SC2_GN_WP1.pdf).
2. Oceanic Fisheries Programme, Secretariat of the Pacific Community, WCPFC-SC2-2006/ST IP-1, 2nd Annual Session of the Western and Central Pacific Fisheries Commission Scientific Committee, Manila, Philippines, 7 to 18 August 2006 (http://www.wcpfc.int/sc2/pdf/SC2_ST_IP1.pdf).
3. At least one major global research initiative, CLIOTOP, has been established to improve understanding of oceanic top predators and their ecosystem (http://web.pml.ac.uk/globec/structure/regional/cliotop/cliotop.htm).
4. N. Daan, V. Christensen, P. Curry, *ICES J. Mar. Sci.* **62** (2005).
5. J. Hinke *et al.*, *Ecol. Soc.* **9**,10 (2004).
6. For example, (*7*) and (*8*). A U.S. National Academies of Science study (*9*) concludes the Myers and Worm (*7*) claim of 90% decline in tuna abundance is an "overestimation."
7. R. Myers, B. Worm, *Nature* **432**, 280 (2003).
8. B. Worm *et al.*, *Science* **314**, 787 (2006).
9. National Research Council, *Dynamic Changes in Marine Ecosystems: Fishing, Food Webs, and Future Options* (National Academies Press, Washington, DC, 2006).

CORRECTIONS AND CLARIFICATIONS

Reports: "A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance" by L. M. Márquez *et al.* (26 Jan., p. 513). On page 514, in the legend to Fig. 4, the colors of the histogram are inverted: The number of dead plants is black, and the number of alive plants is white.

Reports: "From plant traits to plant communities: a statistical mechanistic approach to biodiversity" by B. Shipley *et al.* (3 Nov. 2006, p. 812). In the denominator of Eq. 5 on page 813, the expression in the parentheses should have been set as an exponent of *e*. The correct equation is at right.

$$\hat{p}_i = \frac{e^{\left(\lambda_\infty - \sum_{j=1}^{T} \lambda_{ij} t_{ij}\right)}}{\sum_{i=1}^{S}\left(e^{\left(\lambda_\infty - \sum_{j=1}^{T} \lambda_{ij} t_{ij}\right)}\right)}$$

TECHNICAL COMMENT ABSTRACTS

COMMENT ON "Wetland Sedimentation from Hurricanes Katrina and Rita"

Torbjörn E. Törnqvist, Chris Paola, Gary Parker, Kam-biu Liu, David Mohrig, John M. Holbrook, Robert R. Twilley

Turner *et al.* (Reports, 20 October 2006, p. 449) measured sedimentation from Hurricanes Katrina and Rita in coastal Louisiana and inferred that storm deposition overwhelms direct Mississippi River sediment input. However, their annualized hurricane deposition rate is overestimated, whereas riverine deposition is underestimated by at least an order of magnitude. Their numbers do not provide a credible basis for decisions about coastal restoration.

Full text at www.sciencemag.org/cgi/content/full/316/5822/201b

RESPONSE TO COMMENT ON "Wetland Sedimentation from Hurricanes Katrina and Rita"

R. Eugene Turner, Joseph J. Baustian, Erick M. Swenson, Jennifer S. Spicer

Törnqvist *et al.* accept the usefulness of our data but confuse hurricane landfall location with hurricane storm surge impacts, misrepresent our data interpretation, and misattribute conclusions to our study. Our study did not attempt to address the overall effectiveness of river diversions used for management purposes. We agree that river mouth sedimentation is an important geological process that may lead to marsh colonization.

Full text at www.sciencemag.org/cgi/content/full/316/5822/201c

ENVIRONMENT

How Best to Look Forward?

M. Bruce Beck

In 1999, pandemonium broke out in the normally quiet world of environmental foresight in the Netherlands. Its National Institute for Public Health and the Environment (RIVM), officially charged with preparing the country's State of the Environment reports, was publicly accused of lies, deceit, and shoddy workmanship with its computer models—by one of its own statisticians (*1*). The affair became front-page news, received prime-time coverage on television, and provoked questions and debate in the Dutch parliament. Should geologists Orrin Pilkey and Linda Pilkey-Jarvis enjoy such exalted company as a result of any rumpus arising from *Useless Arithmetic*? I have mixed feelings about that—very mixed, in fact.

I am for the use of quantitative models, which this book is not: not in managing fish stocks across the Grand Banks off Nova

Scotia, in coping with beach erosion and coastal change, in assessing the security of long-term storage facilities for radionuclide wastes under Yucca Mountain in Nevada, nor in any of the other case studies through which the authors focus their assault on such models. Yet, along with the authors, I am allergic to suggestions of our achieving models as truth-generating machines, of virtual realities eliding (in our minds) into realities. Yes, let our quest for explanation be drawn along by this ever-receding horizon; but let us not be deluded into thinking

The reviewer is at the Environmental Informatics and Control Program, Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA 30602–2152, USA. E-mail: mbbeck@uga.edu

we shall ever attain these sunlit uplands.

I am an engineer, moreover, and growing most weary of engineers being the butt of everyone else's jibes, as they are in this book. Yet I know only too well the unquestioning faith placed in models and forecasts by some engineering professionals, about which the authors justifiably complain. I would be as exasperated as they are when the faithful exclaim: "How can my model possibly be wrong, when everything of conceivable relevance has been incorporated into it." This was why Leonard Konikow and John Bredehoeft proclaimed, in the title of their now-famous 1992 paper, "ground water models cannot be validated" (*2*): to jolt and jar the blinkered out of their comfortable routine and to let this be known to the public (for whom the current book has been written). They did not do it, I believe, to place a moratorium on the kind of modeling abhorred by Pilkey and Pilkey-Jarvis—although both paper and book throw down the gauntlet of disproving a negative title, which then is impossible for anyone else to pick up effectively. But since Pilkey and Pilkey-Jarvis are "against" so much in their book, what are they for?

In his novel *State of Fear* (*3*), Michael Crichton offers this perspective on the state of play in modeling and forecasting the impact of climate change on sea levels: "We need more people working in the field, in the actual environment, and fewer people behind computer screens." Pilkey and Pilkey-Jarvis would agree, although not with Crichton's "global warming–trashing novel." I am for data, too: first-class modeling demands first-class data, wherever we can get our hands on them. I wonder, however, at the authors' notion that some definitive, well-chosen fieldwork will alone—free of a model with which to interpret it—eliminate the uncertainty and banish the ambiguity under which we toil. Dealing with the environment is not the stuff of laboratory sci-

Useless Arithmetic
Why Environmental Scientists Can't Predict the Future

by Orrin H. Pilkey and Linda Pilkey-Jarvis

Columbia University Press, New York, 2007. 246 pp. $29.50, £19. ISBN 9780231132121.

ence. Working on aquatic systems akin to Pilkey and Pilkey-Jarvis's case study on abandoned pit mines (their "giant cups of poison"), I have expended substantial funds on acquiring large volumes of high-quality field data. But I have yet to find a model capable of interpreting those data to the point of utter, unambiguous clarity. There are always surprises and conundra; the uncertainty is ineluctable, and not simply because of the inadequacy of models.

How will the authors go forward into the future? For they admire those who attempt to model the propagation of invasive plants through an ecosystem, who possess humility—apparently absent from the likes of me—in the face of staggering complexity. They wish for a "qualitative world" populated by qualitative models, even a "world without [quantitative] models" but with "adaptive management" instead.

My world embraces and celebrates them all. All manner of models, to be chosen (like a set of tools) to suit the particular task. Lay stakeholders in the here and now, doing what they do best (imagining their worst fears and best hopes for the future) with their mental models, fuzzy calculus, or belief network models. Webs of tentative cause-effect couples fed, perhaps, by bizarre boundary and initial conditions, hence to deduce a rich heterogeneity of future patterns of environmental behavior—some radically different from what we have come to know and love. Quantitative modelers (engineers indeed) doing what they do best: mobilizing the current science base (uncertainties, warts, and all) into a computational model; thus to assess the plausibility of the imagined futures, under gross uncertainty; thus to identify what we would most like to understand much better, in the here and now, about the uncertainties (be they elements of the science, the policies, or the technologies) that are crucial to discriminating between the reachability of our worst fears or our greatest hopes. Nothing stands still: not images of the future, nor our qualitative models, nor our quantitative models. Call this living in a "recursive predictive world" [as I have labeled it elsewhere (*4*)], a collective form of adaptive community learning. We condition the step we take today on contemplation of where we might be several steps hence and revisit such conditioning and contemplation each and every day. No hint here of forecasting the state of nature at some point on the globe at some instant in the future.

If, on reading the title of this book, you clench your fist and punch the air while shout-

CREDIT: JOE SUTLIFF

ing "Yes!", you might just stop there. If you wring your hands in despair at the impossibility of picking up the authors' gauntlet, read on and be frustrated, as I was. Be also sore afraid: for someone, somewhere, sometime will make mischief for you with the title.

Today, eight years on from the Dutch "futures" scandal, no one makes more strenuous efforts than does the Netherlands' RIVM to accommodate and cope with the uncertainties of environmental data and models, hence to achieve the greatest possible quality in generating environmental foresight. They had laid the foundations for so doing long before the scandal broke.

References

1. J. P. van der Sluijs, *Futures* **34**, 133 (2002).
2. L. J. Konikow, J. Bredehoeft, *Adv. Water Resour.* **15**, 75 (1992).
3. M. Crichton, *State of Fear* (HarperCollins, New York, 2004).
4. M. B. Beck, Ed., *Environmental Foresight and Models: A Manifesto* (Elsevier, Oxford, 2002).

DANCE

From the Big Bang to the Mind

In a Barcelona theater a few weeks ago, I watched a man and his brain dance simultaneously. Across a big empty stage, the man glided between quadrants defined by a set of glowing crosshairs. Meanwhile, projected onto the wall behind him was an enormous image of a brain, a top-down cutaway view divided by similar crosshairs. Each time the man drifted to a new place on the grid, a splatter of red splotches danced across the brain into a new pattern.

The man was Cesc Gelabert, one of Spain's most celebrated choreographers. The scene was the signature solo performance within a dance called Orion, which he created for the Barcelona festival of science. The brain was indeed Gelabert's, and it was intimately observed by Óscar Vilarroya, a neuroscientist at the Autonomous University of Barcelona, using functional magnetic resonance imaging (fMRI). Long before the performance, Vilarroya loaded Gelabert into an fMRI magnet and played a recording of the accompanying music. The device tracked the evolving pattern of oxygen consumption by Gelabert's hungry brain cells as

he imagined himself going through the motions—and emotions—of his dance.

This collaboration between artist and scientist is a long-standing one. About once a week for more than a year, Gelabert and Vilarroya met for an ongoing conversation. The seed of inspiration was the fact that conscious experience emerges—somehow, astoundingly—from 1.5 kilograms of wet flesh. From that starting point, they rode the reductionist rollercoaster backward in time: If the mind comes from the brain, where does the brain come from? If the brain is the result of a genetic blueprint for wiring up neurons, then where did the neurons and the blueprint come from?

For most of us, the conversation ends when someone notices the light of dawn and sighs, but Gelabert and Vilarroya never dropped the thread. Their discussion of evolutionary biology led to DNA, and thence to chemical evolution on the early Earth. And they kept right on going: from the cooling of our planet to the formation of our Sun; from the aggregation of galaxies to the first atoms, the fundamental particles, cosmic inflation, and finally, of course, the big bang. It is a measure of Gelabert's insatiable curiosity and ambition that he set out to tell the story of this conversation in reverse, in its entirety, through dance.

The performance begins gorgeously. Lights up on seven dancers. As a solo harpsichord arpeggiates out the prelude to Handel's Suite Number Three, the dancers create a zoo of pairwise interactions, encapsulating the behaviors of the fundamental particles. (Don't ask which specific particles are represented by that resounding slap to the buttock of a stony-faced woman— we're not meant to.) With the rules of play suggested, the lights go suddenly dim and bloody as the theater's ably-equipped sound system produces what can only be described as an underground nuclear test played in ultra-slow motion. What makes the ensuing scene work is that we're not allowed to catch our breath. As the dancers bring the rapidly evolving universe into being, they seem to say: "There's too much to tell, so just watch!" The effect is enhanced by a series of minimalist animations, designed by the Spanish filmmaker Paco Periñán, that provide myopic snatches of the sort of trajectory data produced by particle collisions.

Orion
by Cesc Gelabert and the Gelabert-Azzopardi Dance Company
Teatre Lliure, Barcelona, Spain. 22 February – 4 March 2007.
www.gelabertazzopardi.com/pagines/english/orion.htm

From here, we dance through the history of the universe along a logarithmic scale. By the time we reach the midpoint, we've already seen inflation, solar systems, complex molecules, cells, and multicellular organisms, and judging by the sound of bird song in the distance, we're past the Cretaceous by the intermission. The final act is devoted to the past few million years of human evolution, both anatomical and cognitive.

Gelabert is at his best when he keeps the concept abstract, allowing us to focus on the performance. At times he beats us over the head with literalism: During a footrace between knuckle-walking dancers, the losers curl up and play "extinct" while the survivors gradually stand erect. (And just in case you haven't gotten it, Periñán's increasingly distracting projections include pictures of primates, rock art, early weapons, and so forth.) Another source of frustration is that Gelabert never settles on a consistent mode of expression. What begins as dance eventually slides into mime and clowning, such as the delightful conga line of DNA dancers who replicate a series of gestures as mutation sets in. By the end, Gelabert has abandoned dance and relies on props: A woman peels herself out of a full-body condom and collapses in a melodramatic ending that is pure performance art.

These gripes aside, my main complaint about Orion is that it offers too much of a good thing. The dance is divided into thirds: "matter, life, the human being." Once expanded, each of these sections could have formed a successful composition on its own. Gelabert's aim was to capture the thrill and mystery of the scientific worldview. Although the overall effect left me wanting less, not more, he does pull it off in many shining moments.

–**JOHN BOHANNON**

ENVIRONMENT

Environmental Monitoring Network for India

An integrated monitoring system is proposed for India that will monitor terrestrial, coastal, and oceanic environments.

P. V. Sundareshwar,* R. Murtugudde, G. Srinivasan, S. Singh, K. J. Ramesh, R. Ramesh, S. B. Verma, D. Agarwal, D. Baldocchi, C. K. Baru, K. K. Baruah, G. R. Chowdhury, V. K. Dadhwal, C. B. S. Dutt, J. Fuentes, Prabhat K. Gupta, W. W. Hargrove, M. Howard, C. S. Jha, S. Lal, W. K. Michener, A. P. Mitra, J. T. Morris, R. R. Myneni, M. Naja, R. Nemani, R. Purvaja, S. Raha, S. K. Santhana Vanan, M. Sharma, A. Subramaniam, R. Sukumar, R. R. Twilley, P. R. Zimmerman

Understanding the consequences of global environmental change and its mitigation will require an integrated global effort of comprehensive long-term data collection, synthesis, and action (*1*). The last decade has seen a dramatic global increase in the number of networked monitoring sites. For example, FLUXNET is a global collection of >300 micrometeorological terrestrial-flux research sites (see figure, right) that monitor fluxes of CO_2, water vapor, and energy (*2–4*). A similar, albeit sparser, network of ocean observation sites is quantifying the fluxes of greenhouse gases (GHGs) from oceans and their role in the global carbon cycle (*5, 6*). These networks are operated on an ad hoc basis by the scientific community. Although FLUXNET and other observation networks cover diverse vegetation types within a 70°S to 30°N latitude band (*3*) and different oceans (*5, 6*), there are not comprehensive and reliable data from African and Asian regions. Lack of robust scientific data from these regions of the world is a serious impediment to efforts to understand and mitigate impacts of climate and environmental change (*5, 7*).

The Indian subcontinent and the surrounding seas, with more than 1.3 billion people and unique natural resources, have a significant impact on the regional and global environment but lack a comprehensive environmental observation network. Within the government of India, the Department of Science and Technology (DST) has proposed filling this gap by establishing INDOFLUX, a coordinated multidisciplinary environmental monitoring network that integrates terrestrial, coastal, and oceanic environments (see figure, right).

In a workshop held in July 2006 (*8*), a team of scientists from India and the United States developed the overarching objectives for the proposed INDOFLUX. These are to
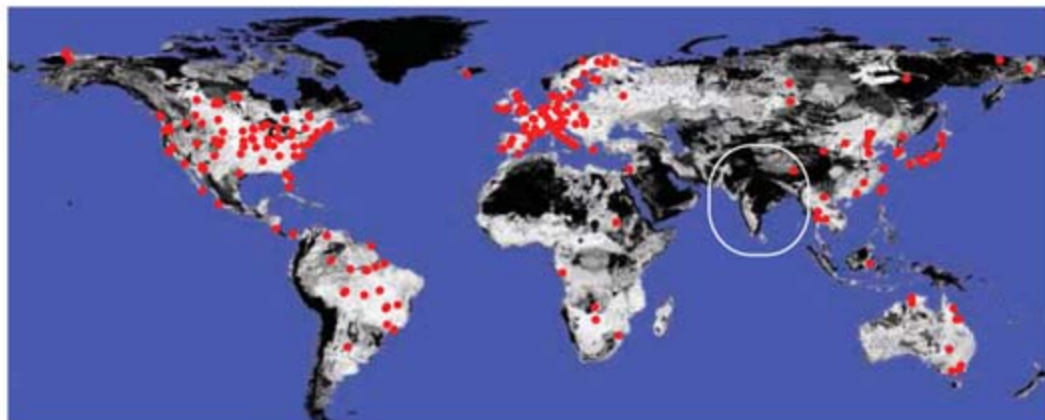
Current monitoring sites in FLUXNET. Sites are shown in red, and global representativeness is estimated by Global Multivariate Clustering Analysis (*24–26*). Darker areas are poorly represented by the existing FLUXNET towers. Environmental similarity was calculated from a set of variables (precipitation, temperature, solar flux, total soil carbon and nitrogen, bulk density, elevation, and compound topographic index) at a resolution of 4 km.

provide a scientific understanding (i) of the coupling of atmospheric, oceanic, and terrestrial environments in India; (ii) of the nature and pace of environmental change in India; and (iii) of subsequent impacts on provision of ecosystem services. Also, in order to evaluate what will enable India to sustain its natural resources, these goals include an assessment of the vulnerability and consequent risks to its social and natural systems.

Climate change will alter the regional biosphere-climate feedbacks and land-ocean coupling. Although global models reliably predict the trend in the impact of climate change on India's forest resources, the magnitude of such change is uncertain (*9*). Similarly, whereas all oceans show the influence of global warming (*10*), the Indian Ocean has shown higher-than-average surface warming, especially during the last five decades (*11, 12*). This warming may have global impacts (*13, 14*), even though the impact on the Indian summer monsoons is not well understood (*15, 16*). These uncertainties highlight the need for regional models driven by regional data.

As the hypoxia observed in the Gulf of Mexico is related to agricultural practices in the watershed (*17*), Indian Ocean studies also indicate couplings between mainland activities and offshore and

A schematic of the INDOFLUX proposal. Placement of stations reflects different climactic, vegetation, and land-use areas. Final locations will be determined as part of the formal science plan.

CREDITS (TOP TO BOTTOM): W. W. HARGROVE/P. V. SUNDARESHWAR (SOURCE)
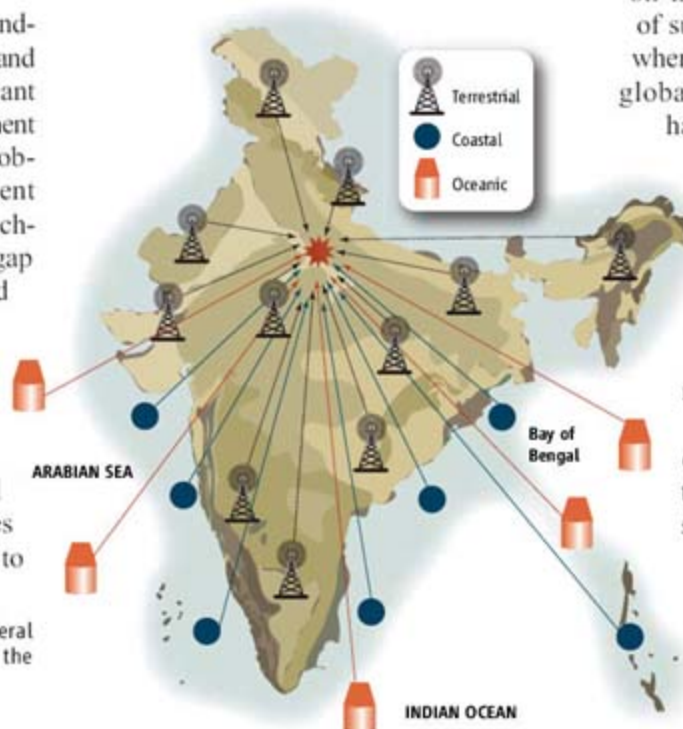
coastal ecosystem processes, with characteristics that differ dramatically between the Arabian Sea and the Bay of Bengal (18). Countries in the Indian subcontinent (India, Pakistan, and Bangladesh) collectively rank among the highest in the ratio of total population to length of coastline, as well as in ratio of land area to length of coast. This suggests that land-use effects on coastal processes and vice versa (i.e., sea-level rise), will be extreme. For developing countries such as India, with a large rural population dependent on the agricultural sector and natural resources, it is imperative to implement scientifically informed policy decisions and management strategies (19).

INDOFLUX will help assess the status of the environment in the Indian Subcontinent and surrounding oceans and will create a baseline from which to evaluate future environmental changes. India's current GHG liability is roughly 1 billion metric tons annually (20). India is presently a non–Annex I country in the Kyoto Protocol of the United Nations Framework Convention on Climate Change and is exempt from binding GHG emissions targets, but as it makes the transition to a developed economy, its status could be revised. India has called for a comprehensive, long-term monitoring network to accurately assess India's GHG inventory and its vulnerability and adaptation to environmental change (20). The proposed INDOFLUX system will provide a map of sources and sinks for carbon and GHGs from various Indian landscapes and will help develop strategies for mitigating domestic and global emissions. In addition to assessing the GHG footprint of ecosystems in the region, the INDOFLUX will also help evaluate the impact of environmental change on ecosystem services provided by them.

The DST has requested about U.S. $50 million for the next 5 years to implement the INDOFLUX network. The money is now earmarked in the 11th 5-year plan. However, for the funds to be released, the scientific community must present a formal plan to the cabinet during the current fiscal year (April 2007 to March 2008).

INDOFLUX will build on ongoing national scientific programs. For example, the India Meteorological Department has initiated a climate-related environmental monitoring program that includes sampling of CO_2 concentrations at various stations. The National Physical Laboratory and their collaborators have assessed the GHG footprint for India (20). The Bose Institute operates a facility at Darjeeling for a comprehensive study of the atmospheric environment of the Eastern Himalayas. Similarly, the Indian Space Research Organization's Geosphere-Biosphere

program collects flux data for GHGs, mainly to understand surface-atmosphere interactions. The National Institute of Oceanography has carried out studies of physical and biogeochemical ocean processes, and the Indian National Centre for Ocean Information Services has been a partner in the Argo program (21) to measure ocean temperature and salinity.

Variables for Indoflux Monitoring

Parameters	Terrestrial	Marine
Net radiation balance	X	X
PAR	X	X
Aerosol optical thickness	X	X
Fluxes of CO_2 heat, energy, & momentum	X	X
Evapotranspiration	X	
Wind speed and direction	X	X
Air temperature	X	X
Land–use change	X	
Precipitation	X	
Soil moisture	X	
Soil temperature	X	
Sea surface temperature		X
Light attenuation		X*
Soil nutrients (C, N, P)	X	
Nutrients (NH_4^+, NO_3^-, PO_4^{3-}, DON, DOP), organic and inorganic C	X	X*
Nutrient loading in runoff		X
Nutrients (Fe, Si)		X*
Atmospheric nitrogen deposition	X	X
Temperature and salinity		X*
Dissolved oxygen		X*
Total suspended solids		X*
Primary production/ standing biomass	X	X
Aerosol deposition	X	X
Biogenic gas flux (CO_2, CH_4, N_2O, DMS)	X	X
Soil respiration	X	
Aquatic respiration		X
Leaf area index	X	

*Profiles in the marine enviornment.

Representative environmental variables to be measured at individual monitoring stations of the INDOFLUX network. Parameters will be finalized as part of the formal science plan. PAR, photosynthetically active radiation.

Most of these efforts occur under the aegis of various organizations and are not sustained; furthermore, the data sets generated are often not integrated. A unique feature of INDOFLUX is the intention to integrate different ecosystem data from the start. Each station in the network will monitor several variables (see table, above). The volume of national-level data generated will require dedicated and secure data warehousing facilities, as well as high-powered computing facilities to convert data into predictive insights through dynamic models.

The sustained observation network should be created in a phased manner under one agency such as the DST, with support from

other departments and ministries. This will immediately serve to increase the return on investments in similar programs by the participating agencies. It will also ensure that INDOFLUX will be a comprehensive network that integrates not only the land, coasts, and oceans but also relies on a multiagency partnership for its operation and governance.

This new program has a critical opportunity to develop protocols, data and metadata standards, and cyber infrastructure to ensure interoperability with other international observing systems projects. Indeed, the validity and global relevance of INDOFLUX and other regional networks is predicated on efficient reciprocal data-sharing and communication protocols with global partners in real time. With the data generated, the scientific community can deliver the necessary scientific foundation for the development of long-term, defensible government policies to tackle environmental change and to meet our obligations under international accords. The data will support economic and strategic cooperation and collective actions internationally.

References and Notes

1. N. Stern, *The Stern Review: The Economics of Climate Change*; www.sternreview.org.uk.
2. D. Baldocchi *et al.*, *Bull. Am. Meteorol. Sci.* **82**, 2415 (2001).
3. R. Valentini *et al.*, *Nature* **404**, 861 (2000).
4. FLUXNET (integrating CO_2 flux measurements), www.fluxnet.oml.gov/fluxnet/siteplan.cfm.
5. R. R. Christian *et al.*, *Estuaries Coasts* **29**, 871 (2006).
6. OceanSITES, Sustained Interdisciplinary Timeseries Environment observation System; http://kela.soest.hawaii.edu/OceanSITES/.
7. C. Huntingford, J. Gash, *Science* **309**, 1789 (2005).
8. INDOFLUX initiative, www.ias.sdsmt.edu/pvs/INDOFLUX.
9. N. H. Ravindranath, N. V. Joshi, R. Sukumar, A. Saxena, *Curr. Sci.* **90**, 354 (2006).
10. S. J. Levitus *et al.*, *Science* **287**, 2225 (2000).
11. C. D. Charles *et al.*, *Science* **277**, 925 (1997).
12. T. R. Knutson *et al.*, *J. Clim.* **19**, 1624 (2006).
13. M. Hoerling, A. Kumar, *Science* **299**, 691 (2003).
14. A. Giannini, R. Saravanan, P. Chang, *Science* **302**, 1027 (2003).
15. R. H. Kripalani, A. Kulkarni, S. S. Sabade, M. L. Khandekar, *Nat. Hazards* **29**, 189 (2003).
16. R. H. Kripalani *et al.*, *Theor. Appl. Climatol.* **87**, 1 (2006).
17. "Gulf of Mexico hypoxia: Land and sea interactions" (R134, Department of Animal Ecology, Iowa State University, Ames, IA, 1999).
18. S. Prasanna Kumar *et al.*, *Geophys. Res. Lett.* **29**, 88 (2002).
19. J. Sathaye *et al.*, *Curr. Sci.* **90**, 314 (2006).
20. *India's National Communication to the United Nations Framework Convention on Climate Change*, (ISBN 81-7371-498-3, Ministry of Environment and Forests, New Delhi, 2004).
21. Argo (ocean monitoring system), www.argo.ucsd.edu.
22. E. Saxon *et al.*, *Ecol. Lett.* **8**, 53 (2005).
23. W. W. Hargrove, F. M. Hoffman, B. E. Law, *Eos* **84**, 529 (2003).
24. W. W. Hargrove, F. M. Hoffman, *Environ. Manage.* **43**, 539 (2004).
25. The INDOFLUX initiative was primarily catalyzed and funded by the Indo–U.S. Science and Technology Forum, in partnership with the Department of Science and Technology (India), the South Dakota School of Mines and Technology, and Anna University. We thank K. Updegraff for comments.

MEDICINE

LIGHT Hits the Liver

Göran K. Hansson

An inflammatory molecule on the surface of certain immune cells increases the concentration in blood of fats associated with atherosclerosis.

Atherosclerosis, the underlying cause of most cases of myocardial infarction, stroke, and gangrene, is the most common lethal disease in Western societies and is expected to become the number one killer globally by 2020 (*1*). It is an inflammatory disease triggered by the accumulation of plasma lipoproteins in the artery wall (*2*). In this scenario, lipids cause inflammation. However, a report by Lo *et al.* on page 285 in this issue (*3*) turns the situation upside-down by showing that two factors produced by immune cells—the cytokines lymphotoxin and LIGHT—cause the amount of lipids in the blood to increase.

Several studies have implicated the tumor necrosis factor superfamily of proinflammatory cytokines in lipid metabolism. Tumor necrosis factor was discovered not only as a soluble protein that induces the death of tumor cells but also as a molecule (cachectin) that causes hypertriglyceridemia and wasting of muscle and fat tissue (*4*). These effects are due to its inhibition of the enzyme lipoprotein lipase, thus limiting the supply of fatty acids for energy production and fat storage. These remarkable metabolic effects of this cytokine did not attract as much attention as its proinflammatory actions and its ability to promote cell death. However, recent findings of tumor necrosis factor secretion from adipose tissue of individuals with metabolic syndrome, a condition predisposing to atherosclerosis, have focused much interest on the metabolic action of this cytokine and its cousins (*5*).

Several of the more than 40 members of the tumor necrosis factor superfamily of proinflammatory molecules are soluble cytokines; others are membrane proteins that can ligate receptors on adjacent cells. There is substantial cross-talk between receptors and ligands. Two members of this family, lymphotoxin and LIGHT, share many features with tumor necrosis factor (the prototypic family member), such as promoting inflammation and host defense against pathogens, and they have been implicated in several inflammatory diseases, including rheumatoid arthritis and Crohn's disease (*6*). Two other superfamily members, CD40 and

The author is in the Department of Medicine and Center for Molecular Medicine, Karolinska Institute, Stockholm, SE-17176, Sweden. E-mail: goran.hansson@ki.se
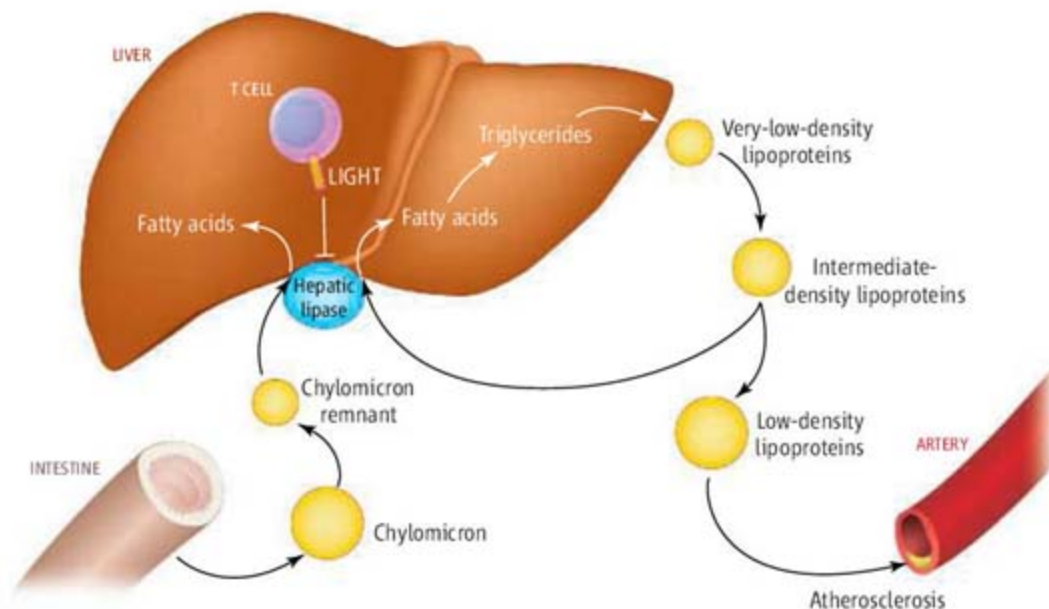
Lipolysis in the liver. Lipoproteins called chylomicrons and very-low-density lipoproteins transport triglycerides and cholesterol in the blood. Hepatic lipase assists in the receptor-mediated uptake of these lipoproteins into the liver. Triglycerides are hydrolyzed, releasing free fatty acids. Remaining very-low-density lipoproteins can be converted into low-density lipoproteins that accumulate in arteries and initiate atherosclerosis. T cells expressing LIGHT inhibit hepatic lipase expression in hepatocytes, which causes increased plasma lipoprotein concentrations, and may contribute to atherosclerosis.

OX40L, are involved in atherosclerosis, and seem to propagate plaque inflammation (*7, 8*).

In contrast to tumor necrosis factor, LIGHT is mainly expressed on the surface of T cells and specialized cells of the immune system called dendritic cells (*9*). Lo *et al.* observed that transgenic mice engineered to overexpress LIGHT on T cells developed hyperlipidemia, displaying elevated cholesterol and triglyceride concentrations in the blood. When T cells from such mice were transferred into normal mice, plasma cholesterol concentration increased substantially. LIGHT (and lymphotoxin) bind to the lymphotoxin β receptor but also to other receptors. By treating mice with soluble forms of the receptor, to function as "decoys," the authors establish that the hyperlipidemic effect of LIGHT depends on the lymphotoxin β receptor.

At first glance, it is remarkable that a cell surface molecule expressed by T cells has such dramatic effects on plasma lipids and lipoproteins. However, other links between cellular immunity and lipid homeostasis have been uncovered. In atherosclerosis, T cells of the T helper I subtype are activated by lipoproteins trapped in plaques on artery walls, thus promoting inflammation (*10*). Natural killer T cells, a subset of T cells, recognize lipids

that are displayed in a complex with CD1 molecules on the surface of antigen-presenting cells (*11*). These T cells accelerate atherosclerosis (*12*) and are enriched in the liver. In the Lo *et al.* study, treatment with soluble lymphotoxin β receptor did not affect plasma cholesterol concentration in mice lacking functional natural killer T cells. This subset of T cells may conceivably regulate lipid homeostasis by delivering LIGHT to the liver (see the figure).

The molecular mechanism by which the lymphotoxin β receptor causes hyperlipidemia was explored by gene-expression arrays. Lo *et al.* found a dramatic, 20-fold decrease in messenger RNA that encodes the enzyme hepatic lipase in the liver of transgenic mice that overexpress LIGHT on T cells. Hepatic lipase activity was also reduced in such mice, but not in transgenic mice that overexpress LIGHT but lack the lymphotoxin β receptor. When normal hepatocytes were exposed to recombinant LIGHT or to T cells overexpressing LIGHT, hepatic lipase expression dropped substantially.

Hepatic lipase is expressed on the surface of hepatocytes in the liver. It promotes the receptor-mediated uptake of plasma lipoproteins that harbor triglycerides and cholesterol and specif-

CREDIT: P. HUEY/SCIENCE

ically catalyzes hydrolysis of the triglycerides (*13*) (see the figure). Both these actions likely contribute to the reduction in the amounts of plasma lipids and lipoproteins observed when LIGHT signaling is blocked by treatment with soluble lymphotoxin β receptor.

The functional role of T cell–dependent control of lipid homeostasis through hepatic lipase is unclear. Perhaps inhibition of lipolysis reduces the amount of energy-rich compounds available to pathogens and/or redistribute fatty acids to other organs during host defense. Studies of infections in mice lacking the LIGHT–lymphotoxin β receptor–hepatic lipase axis may provide interesting answers to these questions.

Lo *et al.* suggest the exciting prospect that increasing hepatic lipase expression

with agents that modulate LIGHT signaling may represent a new therapy for treating of dyslipidemia. It is even possible that inhibition of signaling by the lymphotoxin β receptor could dampen atherosclerosis by improving lipid metabolism as well as by reducing vascular inflammation. In humans, low hepatic lipase activity is associated with increased risk for atherosclerotic heart disease (*14*). However, proatherogenic as well as antiatherogenic effects of hepatic lipase have been observed, both in humans and in experimental animal models, and its biology is not yet fully understood (*13*). Further studies on the metabolic and cardiovascular actions of lymphotoxin, LIGHT, and related family members are awaited with great interest.

References
1. C. J. Murray, A. D. Lopez, *Lancet* **349**, 1436 (1997).
2. G. K. Hansson, *N. Engl. J. Med.* **352**, 1685 (2005).
3. J. C. Lo *et al.*, *Science* **316**, 285 (2007).
4. B. Beutler, A. Cerami, *Nature* **320**, 584 (1986).
5. G. S. Hotamisligil, *Nature* **444**, 860 (2006).
6. T. Hehlgans, K. Pfeffer, *Immunology* **115**, 1 (2005).
7. F. Mach, U. Schönbeck, G. K. Sukhova, E. Atkinson, P. Libby, *Nature* **394**, 200 (1998).
8. X. Wang *et al.*, *Nat. Genet.* **37**, 365 (2005).
9. K. Schneider, K. G. Potter, C. F. Ware, *Immunol. Rev.* **202**, 49 (2004).
10. S. Stemme *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3893 (1995).
11. A. Bendelac, P. B. Savage, L. Teyton, *Annu. Rev. Immunol.* **202**, 49 (2006).
12. E. Tupin *et al.*, *J. Exp. Med.* **199**, 417 (2004).
13. S. Santamarina-Fojo, H. Gonzalez-Navarro, L. Freeman, E. Wagner, Z. Nong, *Arterioscler. Thromb. Vasc. Biol.* **24**, 1750 (2004).
14. K. A. Dugi *et al.*, *Circulation* **104**, 3057 (2001).

CHEMISTRY

Femtosecond Lasers for Molecular Measurements

Ultrafast spectroscopic techniques are finding new applications in molecular detection.

Robert P. Lucht

Chemists and biomedical researchers are now using femtosecond lasers to detect and measure molecules in an increasing number of experiments (*1–4*). This is driven by the commercial availability of reliable laser systems with pulse lengths on the order of 100 fs, fast repetition rates, and high energy in each pulse. Taking advantage of methods for producing a wide range of photon wavelengths, researchers can now expand their use of sophisticated spectroscopic detection techniques. On page 265 of this issue, Pestov *et al.* (*5*) report the detection of *Bacillus subtilis* spores (a surrogate for anthrax). In doing so, the authors have not only targeted a substance of vital interest but have advanced the wider use of femtosecond spectroscopy for rapid and selective detection.

One of the most powerful techniques for molecular detection is coherent anti-Stokes Raman spectroscopy (CARS). In this method, depicted schematically in the figure, two laser pulses (historically called the pump and Stokes pulses) create a coherent excitation of molecules (called the Raman coherence) in the sample at time t_0. A third probe pulse interacts with this coherent state at time t_1, creating a signal (the anti-Stokes pulse) that can be

used to map out the molecular resonances that identify a particular chemical species.

Spectroscopists have used CARS with nanosecond lasers for decades (*6*). Typically, the technique relies on pulsed lasers with repetition rates of 10 Hz. Femtosecond laser systems offer the potential for drastic improvements in the capabilities of CARS diagnostic systems. Data acqusition rates of 1 kHz or greater can be achieved, provided that a CARS signal can be acquired with every laser pulse. Femtosecond laser systems also offer the potential to minimize the nonresonant back-

ground and eliminate the effects of molecular collisions on CARS signal generation.

Pestov *et al.* describe a hybrid CARS technique in which a 50-fs pump and Stokes pulses are used to induce a Raman coherence in a target molecule. A "shaped" probe pulse is used that has a much narrower bandwidth and, consequently, a much longer pulse length. As discussed by Pestov *et al.*, the Fourier-transform-limited pump and Stokes pulses (that is, minimum duration for their spectral bandwidth) are optimal for excitation of the Raman coherence. In terms of probing
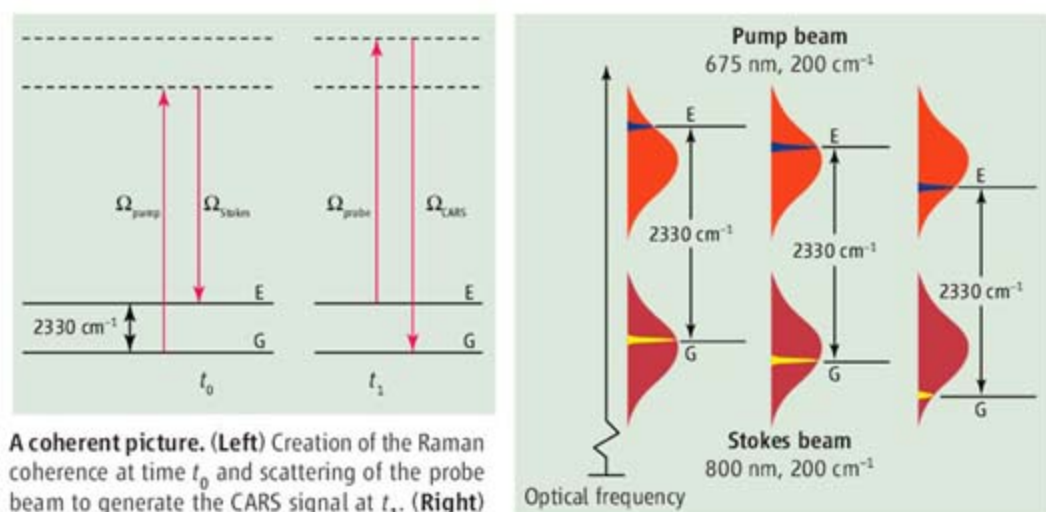
A coherent picture. (Left) Creation of the Raman coherence at time t_0 and scattering of the probe beam to generate the CARS signal at t_1. **(Right)** Schematic illustration of the pump-Stokes frequency pairs that contribute to the excitation of the Raman coherence for Q-branch transitions in the 2330 cm^{-1} vibrational band of N$_2$.

The author is in the School of Mechanical Engineering, Purdue University, 585 Purdue Mall, West Lafayette, IN 47907–2088, USA. E-mail: lucht@purdue.edu

the coherence, however, there are numerous tradeoffs in choosing the spectrum, temporal shape, and time delay of the probe pulse to maximize the resonant CARS signal and minimize the contribution of the unwanted background to the signal. Pestov et al. find that the frequency spectrum of the CARS signal as a function of probe time delay $t_1 - t_0$ contains information vital for selective detection of the molecule of interest.

The excitation of Raman coherences using femtosecond pump and Stokes pulses is surprisingly efficient, provided that the pulses are Fourier-transform-limited. The frequency linewidths of Q-branch transitions (where the change in rotational quantum number $\Delta J = 0$) in the fundamental vibrational Raman band of N_2 near room temperature and pressure are approximately $0.1\ cm^{-1}$, a factor of more than 1000 smaller than the $150\ cm^{-1}$ frequency widths of the 100-fs pump and Stokes beams. Despite this drastic difference in frequency widths, there are numerous pump-Stokes frequency pairs underneath the frequency envelopes of the pump and Stokes beams that are separated by the transition frequency of $2330\ cm^{-1}$ (see the figure). If the pump and Stokes beams are Fourier-transform-limited, then these frequency pairs will all contribute to the excitation of Raman coherence with the same phase thus amplifying the signal. Our calculations (7) show that for pump and Stokes irradiance levels of $5 \times 10^{17}\ W/m^2$, the Raman coherence is within a factor of two of its maximum possible value (8).

Once the Raman coherence is excited by the pump and Stokes beams, the probe beam generates a signal with the desired information. The probe beam can be delayed relative to the pump beams in a controlled manner to monitor the temporal evolution of the CARS signal. This technique has been used to measure, for example, spectroscopic parameters and line-broadening coefficients for nitrogen and other molecules (9) and gas temperatures from the frequency-spread dephasing of the macroscopic Raman coherence (10).

However, the same information can be obtained by chirping the probe pulse (11) and spectrally resolving the resulting CARS signal (12); in this case, the temporal decay information is encoded in the spectrum of the CARS signal. Knutsen et al. (2) used a chirped probe pulse to obtain high spectral resolution in the CARS spectra of organic liquid molecules. Pestov et al. use a shaped probe pulse to discriminate against the nonresonant four-wave mixing interference, and spectrally resolved the CARS signal for different probe delays to take advantage of both the temporal and spectral characteristics of the femtosecond CARS signal.

The impressive work of Pestov et al. illustrates the continued rapid development of femtosecond CARS techniques. But femtosecond laser systems are by no means limited to coherent Raman processes. A diagram similar to the figure can also be drawn for two-photon absorption resonances such as the 1S-2P transition in atomic hydrogen, which is of fundamental interest in atomic physics. Femtosecond lasers are also widely used for two-photon-excited fluorescence microscopy (13). Meshulach and Silberberg (14) used a pulse shaper to demonstrate quantum control of two-photon transitions in cesium. Unlike two-photon absorption experiments with nanosecond laser systems, entire vibrational bands of molecules will be excited when femtosecond-laser systems are used. As with femtosecond CARS, the challenge for researchers will be to take advantage of the characteristics of femtosecond laser radiation to develop new methods for probing these two-photon excited states.

References and Notes

1. C. C. Hayden, D. W. Chandler, *J. Chem. Phys.* **103**, 10465 (1995).
2. K. P. Knutsen, J. C. Johnson, A. E. Miller, P. B. Petersen, R. J. Saykally, *Chem. Phys. Lett.* **387**, 436 (2004).
3. T. Lang, M. Motzkus, H. M. Frey, P. Beaud, *J. Chem. Phys.* **115**, 5418 (2001).
4. J.-X. Cheng, X. S. Xie, *J. Phys. Chem.* B **108**, 827 (2004).
5. D. Pestov et al., *Science* **316**, 265 (2007).
6. A. C. Eckbreth, *Laser Diagnostics for Combustion Temperature and Species, Second Edition* (Gordon and Breach, Amsterdam, 1996).
7. S. Roy, R. P. Lucht, T. A. Reichardt, *J. Chem. Phys.* **116**, 571 (2002).
8. M. Jain, H. Xia, G. Y. Yin, A. J. Merriam, S. E. Harris, *Phys. Rev. Lett.* **77**, 4326 (1996).
9. G. Knopp et al., *J. Raman Spectrosc.* **33**, 861 (2002).
10. A chirped pulse is lengthened, compared to the Fourier-transform limited pulse, and the wavelength changes systematically during the pulse.
11. R. P. Lucht, S. Roy, T. R. Meyer, J. R. Gord, *Appl. Phys. Lett.* **89**, 251112 (2006).
12. T. Lang, M. Motzkus, *J. Opt. Soc. Am.* B **19**, 340 (2002).
13. W. Denk, J. H. Strickler, W. W. Webb, *Science* **248**, 73 (1990).
14. D. Meshulach, Y. Silberberg, *Nature* **396**, 239 (1998).

10.1126/science.1141973

IMMUNOLOGY

The Sources of a Lipid Conundrum

Jerold Chun

Distinct sources of a bioactive lipid are needed for lymphocytes to enter the circulation. Therapeutic strategies need to take into account the existence of more than one lipid reservoir.

From diet to waistline, our cultural preoccupation with avoiding fats obscures the fundamental importance of these and related molecules in our lives. Even rare forms of fat have important physiological consequences. A notable example is the lysophospholipids, low-abundance and structurally simple fat molecules derived from cell membranes. Over the last decade, mechanistic understanding of how these lipids are bioactive has revealed their expansive in vivo biology (1). On page 295 in this issue, Pappu

The author is at the Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. E-mail: jchun@scripps.edu

et al. (2) further develop the story of an immunomodulatory lysophospholipid called sphingosine 1-phosphate (S1P). By investigating the sources of this lipid, a clearer picture of its effects on lymphocyte behavior emerges that may contribute to the development of better immunosuppressive drugs.

The cornerstone for understanding the biological effects of lysophospholipids was discovering that their actions are mediated by cell-surface receptors (3, 4). This indicated that lysophospholipids were not lytic (as the prefix "lyso" originally implied), nor were they detergents or constituents of passive membranes, as had been widely assumed based on their structure. Rather, these lipids appeared more like "keys" that open (activate) receptor "locks." A combination of molecular biology and genetics, on one hand, with lipid biochemistry and pharmacology, on the other, has identified what is now a mammalian family of five S1P and five lysophosphatidic acid (a related lipid) receptors that belong to the well-characterized G protein–coupled receptor superfamily.

Paramount to understanding the in vivo functions of these lipid signals has been the use of "knockout" mice genetically engineered to lack one or more lipid receptors. This approach has proved receptor identity and revealed essential in vivo roles broadly affecting physiology, including the nervous, reproductive, cardiovascular, and immune

systems. Without cloned receptors and knock-out mice, lysophospholipid biology would likely still be mechanistically ambiguous, just as it was a decade ago when the first receptors were identified (4).

Pappu *et al.* have extended this molecular genetic approach to study sphingosine kinase-1 and -2, the enzymes that produce S1P (5). Although S1P is predominantly synthesized within cells, it can be released to act as an extracellular signaling molecule. The authors found that genetic removal of both enzymes from hematopoietic, endothelial (vascular and lymphatic), and liver cells in a mouse effectively eliminated S1P circulating in the blood (plasma) and lymph. The resulting physiological consequences, including compensatory

(accompanied by embryonic lethality), thus demonstrating an essential role for these enzymes in normal S1P production (7). It also raised the possibility of removing the kinases (and therefore S1P) at a desired age and in specific cell types through genetic approaches. By using a special kind of targeted mutation, whereby expression of both sphingosine kinases could be conditionally prevented in hematopoietic cells and vascular and lymphatic endothelial cells, Pappu *et al.* identified red blood cells as a major sink for S1P. This was further strengthened in mice lacking other blood components (platelets or B and T lymphocytes). These animals retained normal blood concentrations of S1P, consistent with different approaches (8) that also identified

relating to how loss of the receptor $S1P_1$ prevents lymphocyte egress. One model, based predominantly on genetic approaches, proposes a direct effect of S1P on $S1P_1$ expressed on lymphocytes (6). However, another model, based on the use of chemical compounds, assigns a key role to $S1P_1$ on blood vessels (endothelial cells) where the receptors are proposed to act as "gatekeepers" for lymphocyte egress (9). The methodological details of studies supporting each model do not allow a clear resolution of discrepancies. However, two factors support the first model. First, most if not all data derived from genetic mutants of S1P signaling are in concert with the lymphocyte $S1P_1$ model. Pappu *et al.* provide further support through reconstitution experiments in mice in which $S1P_1$ expression is removed from lymphocytes while maintained on the endothelium. This disrupts egress, and indicates that, at the very least, $S1P_1$ on lymphocytes matters. Second, a "home-grown" approach to finding specific chemical compounds that activate (agonists) or block (antagonists) S1P receptors is promising, but at a comparatively preliminary stage compared to highly modified and carefully tested compounds that might someday become a medicine (10). Just as proven medicines can have unintended side-effects, so too can preliminary compounds produce results that may be unexpected or at odds with other methodologies.

Distinct sources. Sphingosine 1-phosphate (S1P) that is produced in distinct locations acts on receptors ($S1P_1$) expressed on lymphocytes harbored in different organs and tissues. This controls lymphocyte movement into the circulation (blood or lymph).

up-regulation of lymphocyte S1P receptor expression and a major reduction in lymphocyte (T cell, and to a lesser extent, B cell) number in blood and lymph, could be understood based on knowledge of S1P receptor signaling mechanisms in the immune system. In other words, by removing the cell's key-making machinery (sphingosine kinases), the keys are lost and locks stay shut, as expected. This is an important result because it eliminates nonreceptor mechanisms of S1P action in accounting for a well-characterized aspect of in vivo biology—the control of circulating lymphocyte numbers.

During lymphocyte egress, lymphocytes leave their immunological "motels" (lymph nodes, spleen, and thymus) to circulate in the blood and lymph. Previous work demonstrated that an S1P receptor subtype (designated $S1P_1$) is required for this egress because its genetic removal prevents this movement (6). Much less clear has been the source of S1P that allows this egress. Previous studies in mice in which both sphingosine kinases were deleted in all cells showed a profound loss of S1P

red blood cells as a source of S1P rather than platelets and thrombocytes, long thought to be major S1P sources. The Pappu *et al.* study also refines the contributions of whole-blood components to concentrations of other lysophospholipids (4).

Pappu *et al.* further demonstrate that blood-derived S1P supports the migration of lymphocytes from some locales (spleen and thymus) but not others (lymph nodes). Through clever reconstitution experiments in mice that involved irradiating the animals and replacing their bone marrow (the source of hematopoietic cells), they determined that S1P in the lymph node is from a radiation-insensitive cell source (possibly lymphatic endothelium) that is not in blood. Moreover, this nonblood source also supports lymphocyte migration from the spleen and thymus as well (see the figure). Thus, two distinct sources of S1P—red blood cells and intrinsic radiation-resistant cells present in the lymphocyte "motels"—are needed for normal lymphocyte egress.

Pappu *et al.* address another controversy

The therapeutic future of these and other lipid signaling molecules is bright, yet as with many new approaches, unknowns persist. The finding of large reservoirs of S1P in red blood cells presents challenges to strategies proposed for altering blood S1P concentrations as possible therapies [for example, by intravenous delivery of antibodies against S1P into blood (11)]. Additionally, based on the dual-source data of Pappu *et al.*, altering blood concentrations of S1P may leave local sources of S1P untouched. Other unknowns include as-yet unidentified S1P receptors, interactions with other molecular pathways, clear identification of signaling versus structural lipid pools, and relationships to other lipid pathways, as have been observed for prostaglandins (12). A more advanced therapeutic strategy that has entered human trials is the nonselective S1P receptor agonist known as FTY720 or fingolimod, which is being assessed as an immunosuppressive agent for organ transplantation and multiple sclerosis (13–15). At least part of the mechanism through which this experimental drug works, benefiting further from the work of Pappu *et al.*, is basically conserved between mouse and humans, by virtue of the central role played by S1P receptors. It can be expected that the coming years will bring new

mechanistic and therapeutic insights as these "fat" keys open more locks.

References

1. S. E. Gardell, A. E. Dubin, J. Chun, *Trends Mol. Med.* **12**, 65 (2006).
2. R. Pappu *et al.*, *Science* **316**, 295 (2007); published online 15 March 2007 (10.1126/science.1139221).
3. T. Hla, *Prostaglandins Other Lipid Mediat.* **77**, 197 (2005).
4. I. Ishii, N. Fukushima, X. Ye, J. Chun, *Annu. Rev. Biochem.* **73**, 321 (2004).
5. M. Maceyka, S. G. Payne, S. Milstien, S. Spiegel, *Biochim. Biophys. Acta* **1585**, 193 (2002).
6. M. Matloubian *et al.*, *Nature* **427**, 355 (2004).
7. K. Mizugishi *et al.*, *Mol. Cell. Biol.* **25**, 11113 (2005).
8. P. Hanel, P. Andreani, M. H. Graler, *Faseb J.* (2007).
9. H. Rosen, M. G. Sanna, S. M. Cahalan, P. J. Gonzalez-Cabrera, *Trends Immunol.* **28**, 102 (2005).
10. T. Bartfai, G. V. Lees, *Drug Discovery from Bedside to Wall Street* (Elsevier Academic Press, Burlington, MA, 2006).
11. B. Visentin *et al.*, *Cancer Cell* **9**, 225 (2006).
12. X. Ye *et al.*, *Nature* **435**, 104 (2005).
13. V. Brinkmann, J. G. Cyster, T. Hla, *Am. J. Transplant.* **4**, 1019 (2004).
14. L. Kappos *et al.*, *N. Engl. J. Med.* **355**, 1124 (2006).
15. T. Baumruker, A. Billich, V. Brinkmann, *Expert Opin. Investig. Drugs* **16**, 283 (2007).

10.1126/science.1142239

CHEMISTRY

Putting Order into Polymer Networks

Peter M. Budd

Organic, three-dimensional microporous structures have been synthesized. Such "organic zeolites" are light and chemically versatile, offering a range of possible applications.

Microporous materials contain pores or channels with diameters of less than 2 nm—only a little bigger than many molecules. These pores or channels may be used as filters that allow some species through but not others, as containers to isolate or store specific molecules, or as tiny chemical reactors. Chemists have found ways to prepare a wide variety of porous materials, but it has proved difficult to form organic polymer networks with perfectly controlled pore dimensions—until now. On page 268 of this issue, Yaghi and co-workers report the generation of highly porous, organic, three-dimensional crystalline covalent networks (see the figure) (*1*).

Inorganic microporous materials, such as zeolites, usually have well-defined network structures made up of silicon and aluminum atoms linked via oxygen atoms. Windows and cages within the zeolite framework allow small molecules to access high internal surface areas. Some types of zeolite are found naturally, but many more have been synthesized in the laboratory. In recent years, many researchers have tried to create organic analogs of zeolites. Organic components, although inevitably limited in the temperatures they can withstand, allow much greater control over the chemical nature of the accessible surface. For example, groups may be incorporated that recognize specific molecules or catalyze particular reactions. Furthermore, materials based only on light elements are advantageous in applications where mass must be kept to a minimum, such as for storing hydrogen onboard vehicles.

There is some truth in the saying that nature abhors a vacuum. Thermodynamics works against the formation of tiny spaces and the surfaces that surround them. Zeolites are usually metastable structures and, given appropriate conditions, they will transform to more stable, denser phases. During the preparation of porous materials, the spaces are filled with solvent or other small molecules. All too often, a promising-looking structure disintegrates when the molecules propping up the pores are removed. Many new "porous" materials have been reported, but the porosity is often not permanent in the sense of the material remaining intact on evacuation of the pores.

An important step toward organic zeolites was achieved with the development of materials in which rigid organic components are linked by noncovalent interactions, such as metal-ligand or hydrogen bonds. In particular, crystalline metal-organic frameworks have been produced that exhibit impressive levels of gas uptake (*2*, *3*).

Crystallinity is not necessary for microporosity. Indeed, amorphous porous materi-

The author is at the Organic Materials Innovation Centre, School of Chemistry, University of Manchester, Manchester M13 9PL, UK. E-mail: peter.budd@manchester.ac.uk
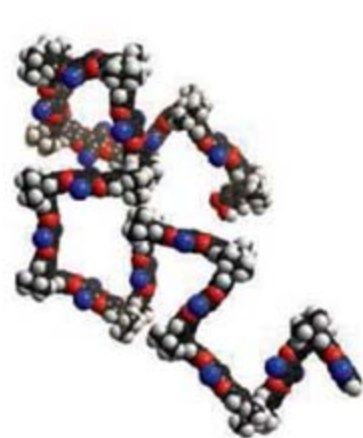
| Membrane-forming microporous polymer | Network-forming microporous polymer | Two-dimensional covalent organic framework | Three-dimensional covalent organic framework |
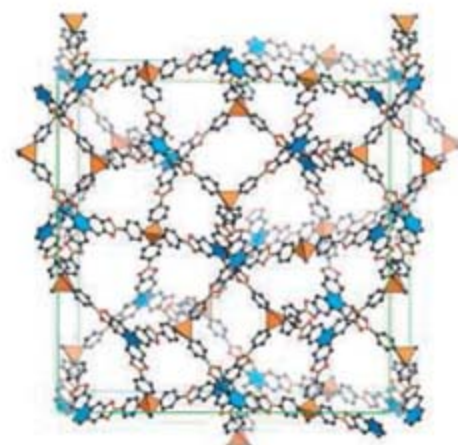
From disorder to three-dimensional order. Porous polymers can display different degrees of order. From left to right, this figure shows a soluble, membrane-forming microporous polymer (*6*), a network-forming microporous polymer that exhibits high gas uptake (*8*), a two-dimensional covalent organic framework (*9*), and one of the three-dimensional covalent organic frameworks described by Yaghi and co-workers in this issue.

als, such as activated carbons, are widely used in industry. In recent years, microporous organic polymers have been created that are chemically well defined, even though they are disordered and therefore possess a distribution of pore sizes.

One approach to microporous polymers involves tying polymer chains together with a large number of rigid bridges, to give "hyper–cross-linked" polymers (4, 5). Another approach starts with the design of the polymer backbone. By connecting rigid ladder-like components with units that force the backbone to twist or turn, it is possible to construct polymers that cannot pack together and fill space efficiently in the solid state. A variety of these "polymers of intrinsic microporosity" (PIMs) have been developed (6–8) (see the figure). Some are soluble and can be processed into useful forms such as membranes, whereas others are three-dimensional networks. They are commonly prepared by making use of a reaction that joins two aromatic rings together with a pair of oxygen bridges. This approach has the potential to generate polymers that are ordered in two or three dimensions, but in practice, amorphous materials are generally obtained.

Current theories suggest that to form a crystalline polymer network, the polymerization reaction must be reversible, so that it occurs under thermodynamic control. Yaghi's group set out to generate ordered polymer networks by making use of reversible condensation reactions of boronic acids. Using this approach, they first produced two-dimensional covalent organic frameworks that incorporate carbon-boron-oxygen linkages (see the figure) (9). A similar, but slightly easier, route to a two-dimensional covalent organic framework was taken by Lavigne and co-workers (10). But extending this concept to three-dimensional covalent organic frameworks is not trivial, because any given combination of building blocks could potentially give rise to an enormous variety of products. In their latest work (1), Yaghi's group drew on their experience of porous frameworks to select the most realistic targets and used a computer model to help predict the structures that were likely to form. This helped them to design the synthesis and identify the products.

The results open a new chapter in the story of porous organic materials. It is likely that routes will now be found to a host of novel crystalline covalent networks. Their high porosity and controllable pore size, coupled with the versatility of organic synthesis, promises that this will be a rich and fruitful area of research.

References
1. H. M. El-Kaderi et al., Science **316**, 268 (2007).
2. A. G. Wong-Foy, A. J. Matzger, O. M. Yaghi, J. Am. Chem. Soc. **128**, 3494 (2006).
3. X. Lin et al., Angew. Chem. Int. Ed. **45**, 7358 (2006).
4. J.-Y. Lee, C. D. Wood, D. Bradshaw, M. J. Rosseinsky, A. I. Cooper, Chem. Commun. **2006**, 2670 (2006).
5. J. Germain, J. Hradil, J. M. J. Frechet, F. Svec, Chem. Mater. **18**, 4430 (2006).
6. P. M. Budd et al., Adv. Mater. **16**, 456 (2004).
7. N. B. McKeown, P. M. Budd, Chem. Soc. Rev. **35**, 675 (2006).
8. B. S. Ghanem et al., Chem. Commun. **2007**, 67 (2007).
9. A. P. Côté et al., Science **310**, 1166 (2005).
10. R. W. Tilford, W. R. Gemmill, H.-C. zur Loye, J. J. Lavigne, Chem. Mater. **18**, 5296 (2006).

PLANETARY SCIENCE

As Tiny Worlds Turn

Rotational force produced by sunlight may help explain the movement of small asteroids, unusual asteroid orbits, and asteroid pairs.

David P. Rubincam and Stephen J. Paddack

Sunlight changes the rotation rate of an asteroid? The idea seems absurd, but on page 272 Lowry et al. (1) and on page 274 Taylor et al. (2) report observations that indicate sunlight is doing just that to the small asteroid 2000 PH5, and Kaasalainen et al. indicate the same is happening on 1862 Apollo (3). The mechanism is the Yarkovsky-O'Keefe-Radzievskii-Paddack effect, mercifully shortened to YORP. With YORP now on a solid foundation, we may be able to understand a number of strange observations involving small spinning asteroids and asteroid binary systems.

The saga of sunlight changing into spin began with Ivan Yarkovsky, a Polish engineer who realized more than a century ago that the infrared radiation escaping a body warmed by sunlight carries off momentum as well as heat (4). Point this heat in the right direction, and it will function like a rocket motor: Each infrared photon escaping the object carries away momentum, thanks to the relationship $p = E/c$, where p is the photon's momentum, E is its
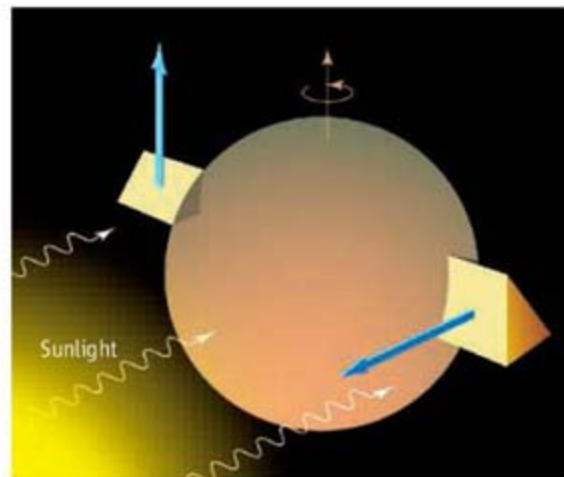
Spinning in the Sun. Sunlight speeds up rotation due to reflection off the vertical and slated faces of the wedges (blue arrows). Infrared radiation emitted by the faces also causes speed-up. If the body spins in the opposite sense, then YORP will slow it down.

energy, and c is the speed of light. By the principle of action-reaction, the object emitting the photon gets a kick in the opposite direction. (Yarkovsky knew nothing of photons and based his reasoning on the outmoded ether concept, but his idea survives the translation to modern physics.) Yarkovsky thrust is tiny, but space is so empty there is no friction to stop it. Moreover, because the Sun is always shining, the Yarkovsky effect goes on century after century with an inexhaustible supply of photonic fuel, profoundly altering the orbits of meter-sized meteoroids (5).

V. V. Radzievskii applied the photon thrust idea to rotation by imagining each face of a cubical meteoroid painted white on one half and black on the other; sunlight reflected by the white part pushes that area more than the black half, causing a torque, which changes the rotation rate (6). His mechanism is weak because the black half, although it reflects little, makes up much of the difference by emitting infrared photons. Moreover, most small solar system objects have fairly uniform albedoes (that is, the fraction of light reflected) across their surfaces.

Building on this work, John A. O'Keefe and one of us (S.J.P.) at NASA realized that shape was a much more effective means of altering a body's spin rate than albedo and set about measuring spin changes in the laboratory. The idea was that light reflecting off of various angled surfaces on the object could

The authors are with NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. E-mail: David.P.Rubincam@nasa.gov

alter its rotation. Thus YORP was born.

O'Keefe and Paddack imagined a body shaped rather like that shown in the figure: two wedges glued to a sphere. As the object rotates, the Sun shines on the vertical face of one wedge and the slanted face of the other. The momentum imparted in the direction of rotation by a photon bouncing off the vertical face is greater than that imparted by bouncing off the slanted face. As a result, there is a net torque, speeding up the object's rotation (7, 8). In addition, infrared radiation emitted by the faces also produces a torque, and infrared YORP probably dominates on small solar system bodies, which tend to be dark. To test this notion, Paddack actually used light (which simulated the Sun) to speed up objects that were magnetically suspended in a vacuum.

Paddack and Rhee applied the YORP effect as the cause of rotational bursting and eventual elimination from the solar system of small asymmetric particles. This may explain the puzzling deficiency in the numbers of small meteoroids (9).

O'Keefe's interest in the YORP effect came from a desire to show that tektites, mysterious glassy lumps found strewn across various regions of Earth, come from the Earth-Moon system (10); farther away, and they spend so long in space that they are spun up to bursting by centrifugal forces via YORP. O'Keefe believed tektites were belched out of lunar volcanoes. Although he was probably wrong

about that—they are most likely created on Earth by giant impacts—and as yet there have been no observations of small celestial bodies being spun up to the bursting point, the YORP effect does have profound consequences for the spins of much bigger objects: asteroids. An object must have a certain sort of lopsidedness for YORP to work, and small asteroids, with their surfaces chiseled by impacts, often have the necessary asymmetry.

A 10-km asteroid might double its rotation rate in a few hundred million years, or have it cut drastically; the asteroid could even end up spinning in the opposite direction (11). Because the YORP time scale is proportional to R^2, where R is the radius, the spin rate of the asteroid PH5 with 57-m radius, already fast with its 12-min period, will double in just 600,000 years. Such rapid time scales argue for YORP being competitive or dominant compared with collisions for changing spin.

In addition to speeding up or slowing down rotation, YORP can alter the axial tilt and precession rate (the rate at which the axis moves) of asteroids (5, 12), so that the entire suite of YORP phenonema can send asteroids into interesting resonant spin states. Moreover, because the strength of Yarkovsky's original effect depends on spin speed and tilt, there should be feedback between orbit and spin evolution.

Most small asteroids are commonly believed to be rubble piles, so that they can change shape or even fission into two smaller

piles as a result of YORP spin-up, helping explain the existence of binary asteroids. And if the body shown in the figure fissions into two wedge-like objects, for example, YORP might continue, so that the binary system orbitally evolves (13). As for further research, is YORP responsible for the Koronis family asteroids having roughly the same spin rate and spin axis orientation (14), or is some other mechanism at work? How precisely does an asteroid change shape or fission as it speeds up? Do asteroids slowed to tumbling ever speed up again? These questions await the future.

References

1. S. C. Lowry et al., Science 316, 272 (2007).
2. P. A. Taylor et al., Science 316, 274 (2007).
3. M. Kaasalainen et al., Nature 446, 420 (2007); 10.1038/nature05614.
4. G. Beekman, J. Hist. Astron. 37, 71 (2006).
5. W. F. Bottke et al., Ann. Rev. Earth Planet. Sci. 34, 157 (2006).
6. V. V. Radzievskii, Dokl. Acad. Nauk SSSR 97(1), 49 (1954).
7. S. J. Paddack, J. Geophys. Res. 74, 4379 (1969).
8. S. J. Paddack, J. W. Rhee, Geophys. Res. Lett. 2, 365, (1975).
9. L. Kresak, in Physics and Dynamics of Meteors, L. Kresak, P. M. Millman, Eds. (Reidel Publishing Co., Dordrecht, Holland, 1968), pp. 391–403.
10. J. A. O'Keefe, Tektites and Their Origin (Elsevier, New York, 1976).
11. D. P. Rubincam, Icarus 148, 2 (2000).
12. V. J. Slabinski, Bull. Am. Astron. Soc. 9, 438 (1977).
13. M. Cuk, J. A. Burns, Icarus 176, 418 (2005).
14. D. Vokrouhlicky et al., Nature 425, 147 (2003).

ECOLOGY

A Positive Feedback with Negative Consequences

The ability to emit isoprene protects some plants from ozone damage, which may affect their future abundance.

Manuel Lerdau

It has long been recognized that isoprene emissions from vascular plants play an important role in atmospheric chemistry (1–3). Recent advances (4, 5) suggest that these emissions—in conjunction with fossil fuel combustion and fertilizer application—may create a positive feedback loop, in which species-specific metabolism causes changes in tropospheric chemistry that in turn affect biological diversity and ecosystem metabolism.

About 1% of the carbon captured during photosynthesis by terrestrial ecosystems is

returned to the atmosphere as isoprene (6). Once in the atmosphere, isoprene can either ameliorate or aggravate ozone pollution, depending on the tropospheric concentrations of reactive nitrogen oxides. When nitrogen oxide concentrations (7) are high, isoprene oxidation leads to the net production of ozone, whereas isoprene oxidation at lower nitrogen oxide concentrations leads to the consumption of ozone (8). High nitrogen oxide concentrations occur in air masses that receive substantial inputs from fossil fuel combustion or fertilizer use. Anthropogenic changes in air chemistry thus alter the atmospheric impact of isoprene.

Isoprene is produced as the first end-product of the methylerythritol phosphate (MEP) pathway, which is responsible for all plastid-derived

isoprenoids in plants (9). Upon synthesis, isoprene diffuses through the chloroplast membrane to the cytoplasm, through the cell membrane to the intercellular air space, and out through stomata (the pores in leaves through which water exits and carbon dioxide enters). Unlike many other organic volatile compounds produced by plants, isoprene is not stored in specialized structures in the plant (10).

We now understand much of the biochemistry of production and the physics of emission. Yet two crucial questions remain: Why do plants make isoprene? And why do only some plants make isoprene? Recent work by Loreto and co-workers attempts to answer these questions (4, 5). In a series of experiments, they exposed leaves to high, but environmentally

The author is at the Blandy Experimental Farm, Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22904, USA. E-mail: mlerdau@virginia.edu

realistic, levels of ozone and also blocked the MEP pathway, thereby restricting isoprene production. The results show that the damaging effects of ozone are ameliorated by isoprene.

The impact of isoprene on the abilities of plants to tolerate oxidative stress is related to the impacts of isoprene on atmospheric oxidant dynamics. Only ~15% of vascular plant species produce isoprene, and most ecosystems consist of both producers and nonproducers. In the northeastern United States, about one-third of tree species produce isoprene; those tropical systems studied have a similar proportion of emitting species (8). Even low-diversity ecosystems such as boreal forests contain a mixture of emitting (spruce and aspen) and nonemitting (pine) species. This heterogeneity creates the possibility for species-specific feedbacks between plants and the atmosphere.

When plants emit isoprene into air with high nitrogen oxide concentrations, these emissions can raise ozone levels, thus increasing the likelihood of oxidative stress damage. However, isoprene-emitting species will tend to be better protected against this damage than those that do not produce isoprene. Those plants that emit isoprene will, in effect, change the atmospheric composition in a way that causes more damage to nonemitters than to emitters (see the figure). Over time, isoprene-producing taxa may become more abundant because of their greater resistance to ozone damage.

This feedback could lead to striking changes in forest ecosystems. For example, in temperate deciduous forests, it provides a mechanism to favor oaks over other trees such as maples, birches, or hickories. In tropical systems, palms and figs will benefit over other groups such as *Cecropia* and mahogany that do not emit isoprene (11, 12).

In recent decades, understanding of the connections between the biosphere and the atmosphere has grown. The influence of biospheric activities on the atmosphere has been seen on global to regional scales. For example, the variation in atmospheric CO_2 concentration throughout the year is a result of seasonal changes in respiration and photosynthesis, and in many ecosystems humidity and rainfall
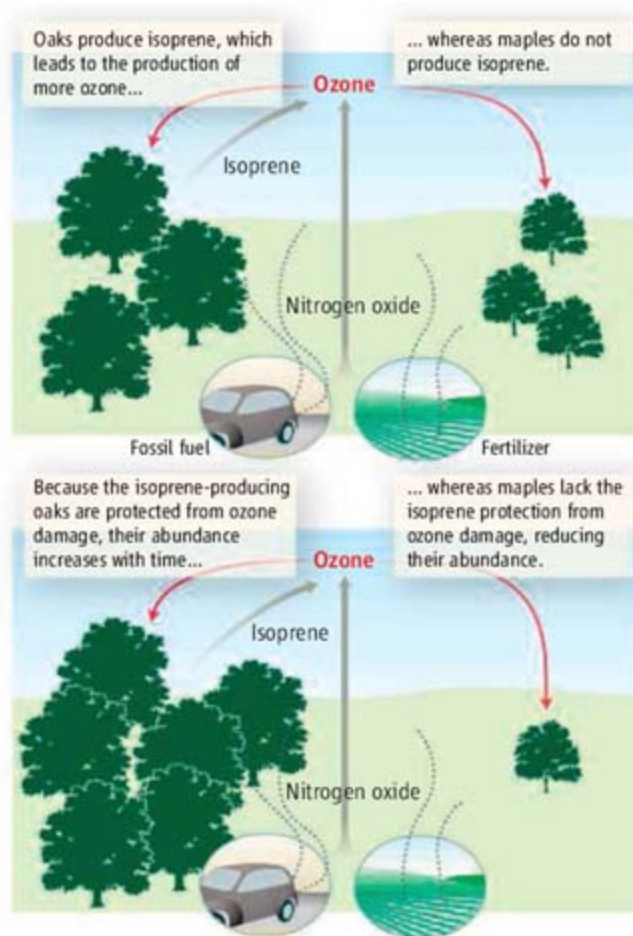
The value of isoprene protection. Interactions between isoprene from an isoprene-emitting species, such as oak, and nitric oxide lead to higher ozone concentrations. This ozone damages species that do not emit isoprene, such as maple, whereas oak trees are protected (**top**). As a result, oak abundance increases, ecosystem-scale isoprene production increases, and ozone concentrations increase, further favoring emitting over nonemitting species (**bottom**).

depend on transpiration from plants. Carbon dioxide concentrations, humidity, and precipitation in turn affect ecosystem metabolism.

To a large extent, however, these feedbacks are not species-specific. Plants of different growth forms, such as trees and grasses, may have different properties, but the metabolisms of different species within a growth form rarely differ enough for such biosphere-atmosphere feedbacks to be important in large-scale models (13). In contrast, the isoprene-atmosphere feedback postulated above operates at the level of species interactions.

The protection from oxidative stress offered by isoprene emission raises the question of why only some plants produce isoprene. Recent molecular phylogenetic work supports the supposition that, at least for flowering plants, isoprene synthase evolved several times independently (14). Plants that produce isoprene tend to tolerate oxidizing conditions better than those that do not, but the production and emission carry a cost in terms of energy and carbon spent and in terms of nitrogen allocated to the enzyme systems (1). The evolution

of isoprene production, or lack thereof, may reflect a balance of the benefits over evolutionary time and the costs such production entails.

It seems unlikely that ozone stress was a major selective force over evolutionary time, because tropospheric nitrogen oxide and ozone concentrations have only become high enough to damage plants since the advent of the internal combustion engine. Isoprene production may have evolved for an entirely different reason: It protects plants against damage from short-term high temperatures (15). The current increase in selective forces favoring isoprene producers—that is, the rise in tropospheric nitrogen oxide concentrations over recent decades (16)—provides an opportunity for isoprene's exaptive (17) value to become important at the scale of communities and ecosystems.

Tropospheric nitrogen oxide concentrations are rising (16) during a time of great changes in vegetation cover, as forest lands are cleared, farmed, abandoned, and then revegetated. Feedback loops such as the one postulated above may push patterns of plant diversity along certain trajectories, and such trajectories may in turn alter the atmosphere. Over time, isoprene-producing species may come to dominate in areas with high nitrogen oxide concentrations. This dominance of isoprene-producing species will decrease plant diversity and have cascading effects on higher trophic levels. As more and more regions become influenced by human activities, the number of communities that suffer in this feedback loop will increase.

References and Notes

1. T. Sharkey, S. Yeh, *Annu. Rev. Plant Physiol. Mol. Biol.* **52**, 407 (2001).
2. G. Sanadze, *Rus. J. Plant Physiol.* **51**, 729 (2004).
3. R. Rasmussen, *Environ. Sci. Technol.* **4**, 667 (1970).
4. F. Loreto, S. Fares, *Plant Physiol.* **143**, 1096 (2007).
5. V. Velikova, F. Loreto, *Plant Cell Environ.* **28**, 318 (2005).
6. A. Guenther et al., *J. Geophys. Res.* **100**, 8873 (1995).
7. A. Fiore et al., *J. Geophys. Res.* **110**, D12303 (2005).
8. C. Wiedinmyer et al., *Earth Interact.* **10**, 10.1175/EI174.1 (2006).
9. H. Lichtenthaler, *Annu. Rev. Plant Physiol. Mol. Biol.* **50**, 47 (1999).
10. R. Monson et al., *Atmos. Environ.* **29**, 2989 (1995).
11. M. Keller, M. Lerdau, *Global Biogeochem. Cycles* **13**, 19 (1999).
12. C. Geron et al., *Atmos. Environ.* **36**, 3793 (2002).
13. P. Friedlingstein et al., *Global Change Biol.* **5**, 755 (1999).
14. T. Sharkey et al., *Plant Physiol.* **137**, 700 (2005).
15. T. Sharkey, E. Singsaas, *Nature* **374**, 769 (1995).
16. L. Horowitz, D. Jacob, *J. Geophys. Res.* **104**, D23823 (1999).
17. Exaptation is the acquisition of a function of a trait other than the one for which that trait originally evolved (18).
18. S. J. Gould, E. S. Vrba, *Paleobiology* **8**, 4 (1982).
19. I thank Z. Cardon, B. Gordon, E. Leger, and H. Mooney for comments on an earlier version. Supported by the Mellon Foundation.

RETROSPECTIVE

Frank Albert Cotton (1930–2007)

Tobin J. Marks

F. A. Cotton was an inorganic chemist who performed groundbreaking work on transition-metal complexes.

On 20 February 2007, the field of chemistry lost one of its most distinguished scholars when Frank Albert Cotton passed away in College Station, Texas, at the age of 76.

Cotton was a world-renowned researcher and educator, and one of a small number of scientists credited with the renaissance of inorganic chemistry that began in the 1950s. His work led to a new appreciation for the diverse structural, bonding, and reactivity properties of transition-metal complexes. As a scholar, collaborator, and mentor, he profoundly influenced the lives of many.

Al Cotton, as he was generally known, was born in Philadelphia on 9 April 1930 and attended local public schools before undergraduate study at Drexel Institute of Technology and then Temple University, both in Philadelphia. During this time, his interest turned from chemical engineering to chemistry, and in 1951, he entered graduate school in chemistry at Harvard University. Here, Cotton studied with the late Sir Geoffrey Wilkinson and was part of the research team that pioneered the synthesis and characterization of ferrocene and related transition-metal organometallic complexes. The nonclassical nature of these complexes dramatically changed the landscape of transition-metal chemistry.

After receiving his Ph.D. from Harvard, Cotton began his academic career at MIT in 1955 as an instructor; in 1961, at age 31, he became the youngest MIT faculty member ever to be promoted to the rank of full professor. In his early years at MIT, Cotton pioneered the application of diverse physicochemical techniques to understanding structure and bonding in transition-metal complexes, explaining previously puzzling physical and chemical phenomena, and discovering new ones. A particularly striking example, published in 1964 in *Science*, was his discovery and analysis of bonding in the first isolable molecule to have a metal-to-metal quadruple bond (*1*). This advance initiated a long and highly productive program in his laboratory to study metal-metal bonding and the properties of metal-metal bonded compounds, from small molecules to metal clusters to extended solids. This interplay among creative synthesis, informative molecular structure, and unusual bonding properties was to continue throughout his career.

Cotton also launched highly productive programs on metal complexes with high coordination numbers (in which the metal ion is surrounded by an unusually large number of bonded ligands) as well as on protein x-ray crystallography. His work on organometallic compounds—which have bonds between metal and carbon atoms—led to fundamental insights into bonding and reactivity. Particularly noteworthy is his discovery and elucidation of the properties of "fluxional" organometallic molecules, in which one or more metal atoms rapidly migrate around the carbon atom framework of the surrounding ligands. The structures of such molecules can be exceedingly nonrigid, and low-energy barrier pathways exist to rapidly interconvert energetically degenerate molecular structures. Cotton elucidated many key aspects of the rates and mechanisms of fluxional processes. The results are relevant to many processes catalyzed by transition-metal ions.

In 1971, Cotton moved to Texas A&M University as Robert A. Welch Professor of Chemistry and was named W. T. Doherty–Welch Distinguished Professor of Chemistry shortly thereafter. Here, his research focused on the synthesis and chemistry of compounds with multiple and/or single metal-metal bonds and other unusual types of structures. His work was recognized with numerous awards, medals, prizes, learned society memberships, and honorary degrees. In 1998, he received the Joseph Priestley Medal, the highest award bestowed by the American Chemical Society. Cotton was scientifically active until his death and published more than 1600 articles in peer-reviewed journals. He served on numerous advisory panels and was a vociferous advocate of support for basic scientific research.

Cotton was also a brilliant scientific educator. He authored or co-authored inorganic chemistry textbooks that set the standard in the field, were translated into as many as 40 languages, and sold more than 500,000 copies. These texts—including *Advanced Inorganic Chemistry*, *Basic Inorganic Chemistry*, *Chemical Applications of Group Theory*, and the high school text *Chemistry, an Investigative Approach*—lucidly convey chemical principles. The same was true of Cotton's scientific lectures, which evidenced infectious enthusiasm, clever wit, and an impressive ability to comprehend and clearly explain complex chemical phenomena.

Cotton was immensely proud of his large family of academic progeny—graduate students (more than 100 Ph.D. recipients), postdoctoral fellows, and their own students—whose accomplishments span many disciplines and continents. As a one-on-one mentor, Cotton had the rare ability to instill in students scientific excitement as well as respect for rigorous scientific methodology, the necessity of hard, careful work, and the lifelong belief that discovery-oriented research is in every way fun. In 2006, Cotton received the George Pimentel Award, the American Chemical Society's highest recognition for achievements in chemical education.

In addition to being a dedicated scholar, Al Cotton was a man who savored life. He was an accomplished equestrian and master of the hunt, a serious gourmet, and a studious wine connoisseur. He loved good music, literature, classical cinema, and travel to exotic places, and spoke several languages fluently. Cotton relished the art of conversation and was an avid debater. He is survived by his wife of 47 years, Diane ("Dee"), and their daughters Jennifer and Jane.

Reference
1. F. A. Cotton *et al.*, *Science* **145**, 1305 (1964).

The author is in the Department of Chemistry, Northwestern University, Evanston, IL 60208, USA. E-mail: t-marks@northwestern.edu

INTRODUCTION

A Barrel of Monkey Genes

THE RHESUS MACAQUE, *MACACA MULATTA*, HAS LONG BEEN USED AS AN important research tool in medicine, from the discovery of the Rh factor to modern-day trials of AIDS vaccines and investigations of neurological and behavioral disorders. As an Old World monkey, it is one of our closest relatives outside of the great apes—chimpanzees, gorillas, and orangutans—making it highly informative on many levels. Today, we unveil the genome of this magnificent Old World monkey and, in doing so, open ourselves to further investigation.

Nonhuman primates have been critically important in developing medical advances that have saved human lives. The availability of this genome sequence will enable new and better experiments that will speed up the pace of research and reduce the number of animals needed for biomedical research in the long run. Analysis of the sequence is stimulating studies on population structure (Hernandez *et al.*, p. 240) that will help researchers interpret differences between experimental groups when they occur and to investigate species-specific effects. Researchers have developed microarray chips of macaque-specific genes so that changes in gene expression in response to perturbations, such as viral infection or hormone treatment, can be followed sensitively through time (see the News article by Pennisi on p. 216). Differences and similarities in the expression of genes related to immune response between macaque and human are already apparent.

Before the rhesus macaque sequencing effort, the human and chimpanzee genomes were available, but these tools could not be used to their fullest potential in studies of evolution. When we look at humans and chimpanzees, we are gazing through a very narrow window, because they diverged so recently in evolutionary history. The rhesus macaque, as our next closest relative, is an ideal outgroup for human/chimpanzee comparisons. We now can step back and examine the past 25 million years of primate evolution, identifying genomic similarities and differences between each species. As shown by the Rhesus Macaque Genome Sequencing and Analysis Consortium (p. 222) and discussed in a second News article by Pennisi (p. 218), the addition of genomic information from macaques is resulting in a much clearer understanding of primate biology and evolution. Within the special section, we also present investigations broadening our understanding of the dynamic nature of primate chromosomes and examining the differences among primate genomes, including the duplications, rearrangements, mobile elements, and evolutionarily new centromeres (Han *et al.*, p. 238; Harris *et al.*, p. 235; and Ventura *et al.*, p. 243). Genes or regions that show evidence of selection among primates may yet provide the Rosetta stone that will allow us to read our own history in the genomes.

–LAURA M. ZAHN, BARBARA R. JASNY, ELIZABETH CULOTTA, ELIZABETH PENNISI

CREDIT: JOSHUA MOGLIA

The Rhesus Macaque Genome

CONTENTS

Science

Breaking new ground. Macaque-specific genomics tools are making studies involving this monkey more useful.

NEWS

Boom Time for Monkey Research

Macaque researchers have blazed a trail of biomedical firsts. Now, with macaque genomic tools at last in hand, this research is rushing ahead in new directions

THE RHESUS MACAQUE IS THE UNSUNG hero of the maternity ward. In 1940, Nobel laureate Karl Landsteiner and his student Alexander Weiner discovered in this monkey a blood protein they called the Rh (for Rhesus) factor. Researchers soon found the Rh factor in some but not all humans and realized that a mother could react immunologically against the factor in her fetus. Now a simple test and a vaccine prevent that reaction—and resulting mental retardation or even death in about 20,000 U.S. newborns a year.

Thanks to Landsteiner, the Rh factor was among the early contributions that this 60-centimeter-tall monkey made to human health. More recently, the macaque has revealed new insights into disorders as diverse as AIDS and depression. But researchers seeking to understand the genetic underpinnings of macaque diseases and behavior have been thwarted. Unlike mice and humans, macaque genetics was virtually unexplored territory until recently, with relatively few genes identified.

And so about 5 years ago, Jeffrey Rogers and Michael Katze decided it was high time

to push macaque biology into the 21st century. Infected by the excitement over the human genome, they decided to go after the macaque genome and to develop the tools to pin down the genes underlying the disorders and behaviors they studied.

They have succeeded in spades. On page 222, they and their colleagues at Baylor College of Medicine in Houston, Texas, describe the high-quality draft sequence of the rhesus macaque (*Macaca mulatta*) genome. Katze, a virologist at the University of Washington, Seattle, used this sequence to develop a macaque-specific microarray that reveals the expression of thou-

Down to the genes. Studies of infant macaque behavior help researchers better understand the genetic basis of shyness.

sands of genes at once. Rogers, a geneticist at the Southwest National Primate Research Center in San Antonio, Texas, has drawn up genetic linkage maps for both baboons and rhesus macaques and is genotyping thousands of macaques in pursuit of specific genes— genes that the sequence will make easier to find. "You now will have the tools and reagents to do in macaque what you can do in humans and in mice," says Katze. "It will completely transform the rhesus as an animal model in human disease in every way."

Already, these efforts are leading to a better understanding of how genes are regulated in diseases such as influenza, and what variants of genes are important for certain behaviors. Researchers are also figuring out how macaques differ genetically from humans—a key step in understanding both the value and limitations of these monkeys as surrogates for humans in experimental work. In AIDS research, for example, it's important to find out where gene expression in humans and macaques diverges. Otherwise, "we are never going to get the right cure or the right [drug] target," says geneticist Timothy Ravasi of the University of California, San Diego.

Building genomic tools

The list of "first discovered in" macaques is impressive. For example, using macaques, researchers demonstrated both that a virus caused polio and, ultimately, the efficacy of the polio vaccine. Since 1985, these animals have been sine qua non for HIV studies and vaccine trials because they can be infected by a simian cousin of the HIV virus that causes a progression of disease similar to that in humans. Then too, monkeys behave much more like people than, say, mice or rats, and they have proven to be good stand-ins for humans in neuroscience and behavioral studies. For example, long-term observations of monkeys raised with their parents or just peers led to key insights into the role of mothering in shaping personalities. Back in the 1990s, the first primate embryonic stem cell came from a macaque. All told, researchers publish about 2000 papers a year on

macaques, with publicly funded researchers conducting studies on about 40,000 animals and drug companies, many more.

But without good genetic tools, macaque research could only go so far. Researchers could document an immune response, for example, but not the changes in gene expression associated with that response. They could identify genetically based behaviors—headbanging similar to that in humans occurs naturally in some macaques—but had no way to track down the genes involved, much less knock out a gene. That's a big contrast with mice, in which researchers have characterized genes involved in dozens of human diseases, thanks to mutagenesis and gene-knockout technologies. Researchers have also developed mouse- and human-specific microarrays—chips or glass slides in which probes of short DNA sequences measure the activity of thousands of genes at once and reveal complex gene circuitries. "I envy people who work in human and mouse," says Shoukhrat Mitalipov, a developmental biologist at the Oregon National Primate Research Center in Beaverton.

Dozens of macaque researchers have made do with microarrays equipped with human DNA probes, but they have never been sure how well the results represented the monkey's gene activity. The average 3% difference between macaque and human genes means that for some genes the macaque sequence may be invisible to a human-based microarray. "If you want to know exactly what's expressed in monkeys, you have to use monkey sequence," says Shrikant Mane, a neuroscientist at Yale University.

Those frustrations drove Katze, Rogers, and their Baylor colleagues to pull together a proposal in 2002 to the National Human Genome Research Institute to sequence the macaque genome. They got the go-ahead in 2005 for the $20 million project, with Baylor's Richard Gibbs leading the sequencing effort and coordinating more than 100 researchers from around the world. As soon as the data started trickling in, Katze teamed up with Agilent Technologies in Santa Clara, California, to put together a macaque microarray based on the new sequences. Several prototypes later, they came up with one with all the macaque's 20,000 genes represented on it.

At the same time, a group led by Robert Norgren, a neuroscientist at the University of Nebraska Medical Center in Omaha, had started on its own gene-chip design. The researchers first used the human DNA sequence to track down the equivalent sequence in macaque, working with Affymetrix Inc. in Santa Clara, California.

Monkey model. Rhesus macaques are the most commonly used primates in biomedical research and are useful in studies from AIDS to depression.

As the macaque genome sequence went online, the chip was expanded to cover all the genes.

Researchers say both macaque-specific microarrays are quite promising. "We can now do comparative genomics at the level of gene expression. [We can ask] how is the macaque genome being expressed and how is it similar or different from the human," says Trey Ideker, a genomicist at the University of California, San Diego.

For Katze, the first task has been to understand how the monkeys react at the genetic level to potentially deadly viral infections. Katze, Yoshihiro Kawaoka of the University of Wisconsin, Madison, and their colleagues infected seven close relatives of the rhesus with a reconstructed version of the flu virus that killed more than 50 million people in the infamous 1918 epidemic; three other monkeys were infected with a modern human flu virus. Three, 6, and 8 days later,

they killed a few of the macaques and analyzed their blood and lungs, using microarrays to study gene expression.

Initially, the macaques' lungs severely overreacted to the 1918 flu virus, the researchers reported in the 18 January issue of *Nature*. With both viruses, the macaques' first line of defense, the innate immune system, kicked in, with genes for inflammatory molecules revving up. In the macaques battling the modern virus, that reaction was temporary, but in those with the 1918 flu, the genes were not only more active but also active much longer, causing extensive tissue damage.

To make matters worse, the subsequent ability of the cell to attack the virus was dampened in the monkeys with the more deadly flu. Type 1 interferon proteins typically activate genes for other proteins that inhibit viral replication. But with the 1918 virus, this genetic pathway seemed disturbingly quiet.

Katze is now one of dozens of infectious-disease researchers using monkey-specific microarrays, including in AIDS research. The microarrays are proving their worth in other disciplines too. To take just one example, neuroendocrinologist Cynthia Bethea of the Oregon National Primate Research Center is using the arrays to delve deeper into the effects of estrogen and progesterone on serotonin, a brain chemical important in mood, appetite, and sex drive.

She and her colleagues have compared gene expression in serotonin-producing nerve cells in menopausal monkeys with and without hormone treatments. Her unpublished results show that with hormone exposure, "there's a dramatic shift" in a biochemical pathway that leads to enhanced production of serotonin, she says. That pathway involves tryptophan, which these nerve cells can use to make either serotonin or a toxin that destroys the nerve cell. In macaque hormone recipients, Bethea's team finds increases in the gene activity of five enzymes used to convert tryptophan to sero-

"As we know much more about the genome, we are in a position to do **much more sophisticated work in this species.**"

—Norman Letvin, Harvard University

tonin and a decrease in five that help produce the toxins. "Our hypothesis is that estrogen and progesterone prevent serotonin neuron cell death and encourage plasticity," Bethea says.

To explore this idea further, Bethea wants to coax embryonic stem cells to become specialized serotonin-producing nerve cells in a lab dish. Here too, the microarrays come in handy, as a tool to examine the nerve cells' gene expression. So far, the chips show that Bethea has some work to do: The lab dish neurons still express many genes typically active only in developing neurons.

Better gene hunts

At the same time that dozens of researchers are building up a picture of overall gene activity in macaques, Rogers has been working toward tracking down specific genes, taking data from many macaque individuals. Last year, he and his colleagues published a genetic linkage map of the rhesus macaque containing known landmarks or bits of identifi-

able DNA, places where the sequence varies from one individual to the next. These maps help researchers home in on specific genes when used with family studies. (The sequence itself, in contrast, comes from a single macaque and provides few clues about what varies between individuals.) And

thanks to long-term breeding programs for the rhesus macaque, Rogers and his colleagues can work with large families whose genealogies are known or can be determined. The stage is set to do genetic epidemiology, says Rogers.

For example, researchers have long studied various behaviors in macaques, including indicators of anxiety or shyness, such as how long it takes an infant to walk away from its mother and explore new surroundings. Judy Cameron of the Oregon primate cen-

Seeing red. A microarray study revealed that although both modern and 1918 flu viruses revved up inflammatory genes early on (red), those genes remained dangerously active in the 1918 virus.

ter has recorded how infants react to such novel situations and found that the exploratory behavior "is strongly heritable," says Rogers.

But until recently, Cameron and others had no way to narrow down where along the macaque's 21 pairs of chromosomes the gene or genes responsible for this behavior are located. Now Rogers and Cameron are using the genetic map to note which landmarks are frequent in infants who are timid or in those who are adventurous, for example. The landmarks will help researchers identify the general vicinity of the genes. Then the team will search the genome sequence at that location for possibly relevant genes and test them. "We're very excited," says Rogers. "This will provide important new information about the genetics of susceptibility to psychiatric disorders among humans."

Other groups are taking a similar tack to uncover genes for vulnerability to stress or risks for neurological and eventually heart and other diseases. Their analyses are just the beginning of a revolution in macaque research. "As we know much more about the genome, we are in a position to do much more sophisticated work in this species," says immunologist Norman Letvin of Harvard Medical School in Boston. "There will be a great deal of work going forward now that these tools are available." Landsteiner would be proud.   **–ELIZABETH PENNISI**

NEWS

Genomicists Tackle The Primate Tree

Primates are taking center stage in genomics, with the macaque serving as an early milestone in understanding our relatives' genomes—and therefore our own

The deciphering of the human genome was a humbling experience. The promise of the project, in the words of James Watson, was "to find out what being human is." But even when most of the 3 billion bases of the human genome had been properly placed, much about the sequence defied understanding. Where in the 20,000 human genes uncovered are the ones that set *Homo sapiens* apart from other mammals, or other primates? To find out, genomicists have been scrambling for more data ever since, most recently from primates. "The goal is to reconstruct the history of every gene in the human genome," says Evan Eichler, a geneticist at the University of Washington, Seattle. And that requires data from our relatives.

DNA from different branches of the primate tree will allow us "to trace back the evolutionary changes that occurred at various time points, leading from the common ancestors of the primate clade to *Homo sapiens*," says Bruce Lahn, a human geneticist at the University of Chicago in Illinois.

In 2005, the unraveling of the chimp genome provided tantalizing hints about differences between us and our closest relative (*Science*, 2 September 2005, p. 1468). Now on page 222, the third primate genome, that of the rhesus macaque, begins to put the chimp and human genomes into perspective. Macaques are Old World monkeys, which split perhaps 25 million years ago from the

ape lineage that led to both chimpanzees and humans (see diagram, p. 219). So when compared to apes, monkeys can help identify the more primitive genetic variants, allowing researchers to tease out the changes that evolved only in apes. Researchers want to take such analyses back to even more ancient evolutionary divergences, and so seven more primate genome sequences are under way, as is the sequencing of the DNA of two close nonprimate relatives. Together, these genomes "should teach us general principles of primate evolution," says Lahn.

A consortium of more than 100 researchers who have been unraveling the macaque genome are detecting genes that have changed

faster than expected in the chimp and human lineages; such speed is usually a telltale sign of significance in evolution. They are also finding that dozens of base changes known to put humans at risk for disease also exist in the healthy macaque—but not in the chimpanzee. That suggests that some gene variants implicated in disease are relics of the ancestral primate condition. Such studies "may be the bridge between comparative genomics and evolutionary biology," says Richard Gibbs, director of the Baylor College of Medicine Human Genome Sequencing Center in Houston, Texas, and coordinator of the rhesus macaque genome project.

Gibbs and his colleagues are tackling evolutionary biology in reverse. They are identifying key genomic differences without yet knowing how or whether those differences translate into traits that provide survival advantages. Traditionally, researchers have first traced changes in the shapes and sizes of beaks, bodies, brains, and so on, then sought the genes behind them. The hope is that the two modes of inquiry will meet in the middle. But so far researchers have come up short in linking genomic changes to traits subjected to natural selection and other evolutionary forces, ironically because of sparse biological data on nonhuman primates, says glycobiologist Ajit Varki of the University of California, San Diego: "[Without] basic information about the chimp, its physiology, its diseases, its anatomy, you are really very impoverished about what you can say."

Beyond mouse

In 2001, the human genome sequence drove home how little we knew about our genomic selves. About one-third of our genes were complete unknowns. Researchers immediately started lining up our DNA with that of worms, fish, and rodents to see what genes matched up and to try to pin down functions. They found not just genes but also conserved regions within the "junk" DNA that played as critical a role in genome function as the genes themselves. Their finds led to an unquenchable thirst for sequence data as a way to clarify how genomes work. "Every additional species increases our ability to resolve functional from nonfunctional [DNA]," explains Ross Hardison, a molecular biologist at Pennsylvania State University in State College.

The surprise of the chimp genome, the first nonhuman primate to be sequenced, was the large number of insertions and deletions that differed between humans and our closest living cousins. There were more changes in the
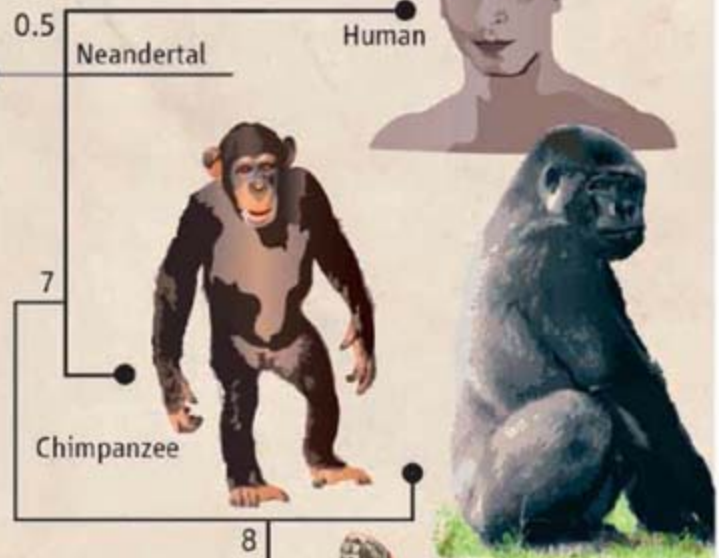
Complete coverage. Researchers plan to eventually sequence the genomes of all these primates and related species, with human, macaque, and chimp now published. The animals are arranged in an artist's rendition of their family tree, with estimated divergence dates in millions of years.
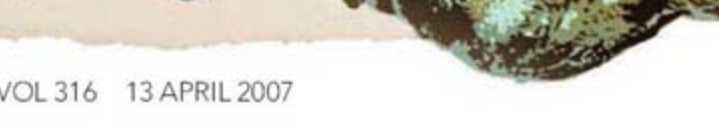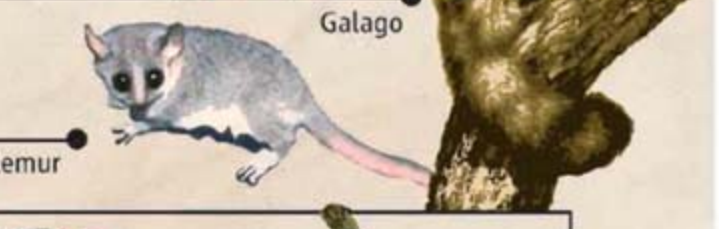
The Rhesus Macaque Genome

order and number of genes and blocks of genes than changes in single base pairs, highlighting the importance of this kind of expansion and shuffling in primate speciation.

But the chimp data proved frustrating as well, because researchers couldn't put the chimp-human comparisons into an evolutionary context. If humans had one base, say a C, at a position where chimpanzees had a G, researchers had no way of knowing which base represented the ancestral condition. Consequently, there was no way to tell whether the change at that position had occurred only in humans—and therefore perhaps helped define *Homo sapiens*—or in the chimp. And so in 2005, the National Human Genome Research Institute began stuffing more primates into the sequencing pipeline and approved the $20 million rhesus macaque sequencing project. "It is great to finally have a [distant relative] that allows us to assign differences between the human and chimpanzee genomes to either the human or the chimpanzee evolutionary lineage," says Svante Pääbo of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany.

Sequencers aren't stopping at the macaque. Sequencing of many primates, including the orangutan, the gibbon, and a New World monkey, the marmoset, is under way, with promises that the baboon should be next. In 2006, the Wellcome Trust Sanger Institute in Hinxton, U.K., started deciphering the gorilla genome, planning coverage similar to that of the macaque. Meanwhile, genomicists have started sequencing key genes and regulatory regions from other primates, too. "To tell what is human-specific, you need this comparative context," says Anne D. Yoder, an evolutionary biologist at the Duke Lemur Center in Durham, North Carolina.

To know a genome

Already, the primate genomic data are revealing bits of our genetic history. For example, more than 98% of chimp and human bases agree. So researchers hoping to pick out areas with fewer base changes than expected—such as regulatory regions conserved in all apes—are awash in a tide of virtually identical DNA. But when the search is expanded to additional primates, there's more variation in the sequences, and previously undetectable conserved regions, even small regulatory

sequences, begin to surface.

For example, Dario Boffelli, now at Children's Hospital Research Center in Oakland, California, and his colleagues at the Joint Genome Institute in Walnut Creek, California, wanted to understand the regulation of genes that help maintain healthy levels of cholesterol in the body. They looked at 558,000 bases covering genes involved in cholesterol processing,

DNA reader. Baylor's Richard Gibbs led more than 100 researchers in the sequencing of the macaque genome.

comparing human and six other primates: baboon, colobus monkey, dusky titi, marmoset, owl monkey, and squirrel monkey. They discovered regions with virtually the same sequence in all the primates. Subsequent experiments showed that three of the newly identified conserved regions do indeed regulate genes in the cholesterol pathway, Boffelli's team reported in January in *Genome Biology*.

In other cases, particularly as researchers look for differences that reflect independent evolution, data from even one additional primate can help. In one analysis, the macaque team looked at 64,000 places in the macaque genome where they knew a disease-related mutation existed. In the past, researchers have assumed that such mutations were specific to humans. A few chimp genes had hinted that some problematic bases might predate humans, but the macaque drives home how often this may be the case. Hardison and his Pennsylvania State colleague Webb Miller found more than 200 sites where the macaque had the same base at the same position as the diseased or at-risk human. In

"**Every additional species increases** our ability to resolve functional from nonfunctional [DNA]."

—Ross Hardison, Pennsylvania State University

97 instances, both the chimp and the macaque matched the aberrant human base; in 48 cases it was just the chimp. And in 84 cases the rhesus, but not the chimp, matched the diseased human sequence, possibly because chimps also independently evolved away from the ancestral condition at those sites.

For example, about 1 in 15,000 people have phenylketonuria because their gene for an enzyme needed to process the amino acid phenylalanine is defective. Untreated, the buildup of a toxic byproduct causes mental retardation. In macaques, that same defective gene is the normal condition and has no ill effects. It could be that many "disease" variants in humans are simply ancestral variants "where [a dietary or environmental] change between the human ancestor and the human has made a variant that used to be good, bad," says Miller.

In addition, the macaque genome consortium combed the macaque, chimp, and human genomes for families of genes that had expanded in one or more species. A family consists of the original gene and any subsequent copies, many of which evolve slightly different sequences and functions over time.

One in particular intrigued Miller. This family, called PRAME—short for "preferentially expressed antigen of melanoma" because the genes are activated in melanoma and other types of tumors—has had a complex history in humans. It has at least 26 intact members on chromosome 1. It's one of the regions of the human genome that "are going wild," says Miller. The chimp has a similarly complex set of PRAME genes, but Miller found just eight PRAME genes in the macaque. "The cluster is very simple [and has] remained stable for millions of years," he explains. Working from this simpler, presumably ancestral set, he and his colleagues hope to unravel the timing and types of duplications that resulted in the abundance of human PRAME genes.

Elsewhere in the genome, the consortium found that the macaque has as many as 33 major histocompatibility complex (HLA) genes, more than triple the number in humans.

CREDIT: RICHARD GIBBS

"When you see a dramatic change, it suggests there was some evolutionary selection that favored those extra copies," says James Sikela, a computational biologist at the University of Colorado Health Sciences Center in Denver. "The tough question is, 'What favored that event?'"

While Sikela and colleagues ponder the macaque's need for HLA genes, Adam Siepel of Cornell University and his collaborators found other genes in which mutations were apparently favored by selection. Such positive selection, as it is called, typically shows up as bases that have mutated faster than would occur by chance. So Siepel's team compared 10,376 macaque genes with their equivalents in both the chimp and human genomes. They sought genes with a relative mutation rate that was higher in bases that changed the encoded amino acid than in bases that did not alter the coding. The researchers found 178 such genes, "considerably more" than previously identified in human-chimp scans, says Siepel. Some genes, such as a few involved in the formation of hair shafts, were changing rapidly in the three species, possibly because climate change or mate-selection strategies spurred rapid evolution, Siepel speculates.

Other positively selected genes detected in at least one species included those involved in cell adhesion and cell signaling, as well as genes coding for membrane proteins. "We don't really know enough at this stage to point to a case where we have a really nice story of a difference at the molecular level that we can connect to a known phenotypic difference," Siepel laments.

"At the genomic level, **evolution is extremely messy,** involving every conceivable mechanism, probably with lots of blind alleys and red herrings."

—Ajit Varki, UC San Diego

Siepel and others say that such stories require more primate sequences. Evidence of positive selection in the same genes in multiple species will provide more clues to what prompted such rapid evolution. Moreover, researchers can be more confident about labeling a gene as "human-specific" once they have looked in a number of our relatives and not found it. "The more primates one can compare, the better," says Sikela.

Sequencing decisions require tough choices about what species to sequence and how thoroughly, however. For his part, Boffelli thinks seven or eight primates would suffice and favors apes over prosimians, the most primitive living primates. With ape DNA, it will be easier to look for positive selection that led to humans. But Yoder thinks it's also impor-

tant to understand how the whole primate branch has evolved, a point long made by researchers studying anatomy and behavior. "If you are going to understand which genes are primate-specific, you need a pretty broad phylogenetic spectrum, [with] things outside the primate clade but close to it," she notes. That argument has already brought tree shrews and flying lemurs (which are not lemurs at all) into the picture, with researchers planning a quick skim across the DNA to get a very rough draft sequence.

Others warn that the quick skim, which is also planned for the bushbaby, mouse lemur, and tarsier, might not be enough, however. With anything short of finished sequence, the computer programs may pick up differences—signs of evolution—that in reality may be sequencing errors, warns Miller. That was the lesson of the chimp genome, which initially was not a very polished draft.

Varki says the genomic work promises to be challenging in other ways, too: "At the genomic level, evolution is extremely messy, involving every conceivable mechanism, probably with lots of blind alleys and red herrings. Deciphering the significance of these molecular changes will be far, far more complicated than I imagined." Nonetheless, Siepel predicts, "we're going to learn a lot in the next 5 years."

–ELIZABETH PENNISI

Know thy genes. The genomes of the gorilla, chimp, orangutan and human (*left to right*) will help clarify our evolution.

CREDITS (LEFT TO RIGHT): JUPITER IMAGES; ANUP SHAH/GETTY (MIDDLE PHOTOS); GETTY

RESEARCH ARTICLE

Evolutionary and Biomedical Insights from the Rhesus Macaque Genome

Rhesus Macaque Genome Sequencing and Analysis Consortium*†

The rhesus macaque (*Macaca mulatta*) is an abundant primate species that diverged from the ancestors of *Homo sapiens* about 25 million years ago. Because they are genetically and physiologically similar to humans, rhesus monkeys are the most widely used nonhuman primate in basic and applied biomedical research. We determined the genome sequence of an Indian-origin *Macaca mulatta* female and compared the data with chimpanzees and humans to reveal the structure of ancestral primate genomes and to identify evidence for positive selection and lineage-specific expansions and contractions of gene families. A comparison of sequences from individual animals was used to investigate their underlying genetic diversity. The complete description of the macaque genome blueprint enhances the utility of this animal model for biomedical research and improves our understanding of the basic biology of the species.

Rhesus macaques (*Macaca mulatta*) (*1*) are one of the most frequently encountered and thoroughly studied of all nonhuman primates (table S1.1). They have a broad geographic distribution that reaches from Afghanistan and India across Asia to the Chinese shore of the Pacific Ocean. As an Old World monkey (superfamily Cercopithecoidea, family Cercopithecidae), this species is closely related to humans and shares a last common ancestor from about 25 million years ago (Mya) (*2*). The two species often live in close association, and macaques exhibit complex and intensely social behavioral repertoires.

The relationship between humans and macaques is even more important because biomedical research has come to depend on these primates as animal models. Compared with rodents, which are separated from humans by more than 70 million years (*2*, *3*), macaques exhibit greater similarity to human physiology, neurobiology, and susceptibility to infectious and metabolic diseases. Critical progress in biomedicine attributed to macaques includes the identification of the "rhesus factor" blood groups and advances in neuroanatomy and neurophysiology. Most important, their response to infectious agents related to human pathogens, including simian immunodeficiency virus and influenza, has made macaques the preferred model for vaccine development. Lesser-known contributions of these animals include their early use in the U.S. space program—a rhesus monkey was launched into space more than a dozen years before any chimpanzee.

The cynomolgus macaque (*M. fascicularis*), pigtailed macaque (*M. nemestrina*), and Japanese macaque (*M. fuscata*) have all contributed to research, but the rhesus macaque has been used most widely. Taxonomists recognize six *M. mulatta* subspecies (*1*), which differ substantially in their geographical range, body size, and a variety of morphological, physiological, and behavioral characteristics. North American research colonies include animals representing both Indian and Chinese subspecies, although India ended the exportation of these animals in the 1970s.

With the advent of whole-genome sequencing, a highly accurate human genome sequence and a draft of the chimpanzee genome have been generated and compared. The chimpanzee shared a common ancestor with humans approximately 6 Mya (*4*, *5*), and the major impact of the chimpanzee genome sequence data has been in their direct comparison with data from the human genome. However, the chimpanzee data have major limitations. First, because the alignable sequence is only 1 to 2% different from that of the human, there is no informative "signal" to distinguish conserved elements from the overall high background level of conservation. This is exacerbated by the fact that the chimpanzee genome was an incomplete draft, containing sequence errors that could potentially mask true divergence. Second, the differences that are found between humans and chimpanzees are difficult to assign as specific to either the chimpanzee or the human. As a result, the chimpanzee analyses have on their own provided relatively few answers to the fundamental question of the nature of the specific molecular changes that make us human.

By contrast, the genome of the rhesus macaque has diverged farther from our own, with an average human-macaque sequence identity of ~93%. Figure 1 shows the inferred common ancestor for all three species, as well as a common ancestor that predated the human-chimpanzee divergence. A characteristic that is found in humans but not in the chimpanzee can be rec-

ognized as a loss in the chimpanzee if it is present in the macaque, or it can be recognized as a gain in the human if it is absent in macaque. In principle, this three-way comparison should make it possible to pinpoint many changes and identify specific underlying mutational mechanisms, which could have been critically important during the past 25 million years in shaping the biology of the three primate species.

We examined the basic elements of the rhesus macaque genome and undertook reconstruction of the major changes in the human-chimpanzee–rhesus macaque (HCR) trio. The regions of the genome that were duplicated in macaque were then identified and correlated with other genome features. Individual macaque genes were studied, and the orthologous genes in the HCR trio were aligned to reveal evidence for the action of selection on individual loci. Additional animals from other populations were also sampled by DNA sequencing to study their genetic diversity. Throughout, complementary methods were applied and the different results combined in order to represent the most complete picture of macaque biology. For a visual representation of some of the insights gained from the genome and more information about the importance of the macaque as a model organism, see the poster in this issue (*6*).

Sequencing the Genome

To generate a draft genome sequence for the rhesus macaque, whole-genome shotgun sequences were assembled. The bulk of the sequencing used DNA from a single *M. mulatta* female, whereas DNA from an unrelated male was used to construct a bacterial artificial chromosome (BAC) library to provide BAC end sequences and to aid in selective finishing. We
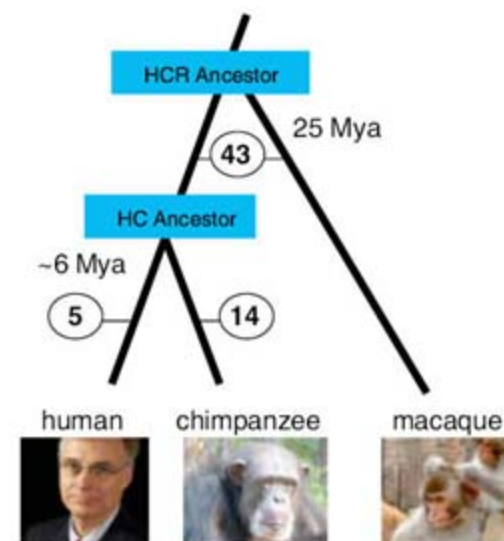
Fig. 1. Evolutionary triangulation in the human, chimpanzee and rhesus macaque lineages (lineage-specific breaks), showing a summary of chromosomal breakpoints on a microscopic scale (Fig. 3) (*7*). Circled numbers indicate numbers of lineage-specific breaks.

*To whom correspondence should be addressed. Richard A. Gibbs, E-mail: agibbs@bcm.edu
†All authors with their contributions and affiliations appear at the end of this paper.

used several whole-genome shotgun libraries with different insert sizes (~3.0, 10, 35, and 180 kb) to generate a total of 18.4 Gb of raw DNA sequence through standard fluorescent Sanger sequencing technologies. Initial assemblies to the intermediate scaffold stage were carried out by the three different assembly methods: Atlas–whole-genome shotgun, parallel contig assembly program (PCAP), and the Celera Assembler (7). These were compared by means of more than 200 metrics, including gross sequence statistics, agreement with finished sequence, utility for gene predictions in the Ensembl pipeline, and accuracy of alignment to the human genome. The three unpolished assemblies were found to be largely similar and of high quality, so all were used in combination with other genome data for the subsequent assembly and placement of long sequence segments on the macaque chromosomes (tables S2.1 to S2.4).

To produce an optimal representation of the genome, the three intermediate assemblies were merged (Fig. 2). Melding the assemblies involved mapping the Atlas–whole-genome shotgun and PCAP data to the Celera Assembler output, which had longer contiguity than the other two data sets at this stage of the process. There was little difference between assemblies at the sequence contig level, at which robust sequence alignments guide the reconstructions, so we focused our attention instead on contigs that were joined into scaffolds. Additional pairs of Celera Assembler scaffolds were joined based on their mapping to the other two macaque assemblies. Analysis of the output showed that this composite assembly was superior to any of its components (table S2.4).

During assembly, a comparison with the human genome sequence [National Center for Biotechnology Information (NCBI) accession code b1d35] identified a small number (<100) of obvious inconsistencies, such as improper joins of different chromosomes. These scaffolds were therefore split at the misassembly

point. The human map was also used to help place large merged scaffolds onto the macaque chromosomes (8, 9) [the chromosome numbering of Rogers et al. (8) was used] at the highest level of the assembly process. Given that the human data were only used to split scaffolds and that de novo macaque assemblies were always given precedence over the mapping to the human genome in the macaque assembly merging and chromosome assignment process, the final product should not be regarded as a "humanized assembly."

The total length of the combined genome assembly was approximately 2.87 Gb (Table 1). This incorporated ~14.9 Gb of raw sequence, which represents about a 5.2-fold coverage of the macaque genome. Comparison with expressed sequence tag (EST) sequence data and approximately 1.8 Mb of finished sequence (see "Selected sequence finishing," below) indicated that ~98% of the available genome was represented. No misassemblies were identified in that comparison. Contigs showed an N50 (minimum length of contigs representing half of the total length of the assembly) of >25 kb; the N50 for sequence scaffolds was >24 Mb. GenBank accession codes are available online (table S2.5).

Selected sequence finishing. The rhesus macaque genome assembly is a draft DNA sequence, and it contains many gaps. A higher data quality with greater contiguity was desired at several genomic regions that attracted additional interest. In these cases, individual BAC clones were isolated, and data quality was improved by sequence "finishing." Many of these BACs were in regions of pronounced genome duplication, whereas others were gene-rich. All finished BACs, their gene content, and their genome coordinates are listed in table S2.6.

Overview of Genome Features

General organization and content. The macaque genome is organized into 20 autosomes

and the XY sex chromosomes. With the exception of 48 breakpoints (Fig. 1)—including three fusions, one fission, and breakpoints induced by inversions that are each detectable through chromosome staining, by radiation hybrid mapping, or by comparative linkage mapping—there is a superficial similarity between the macaque and human chromosomes (8–11). Several chromosomes in the macaque are also more acrocentric than their human counterparts, but many from the two species are difficult to distinguish.

Nucleotide sequences that aligned between the human and rhesus average 93.54% identity. If, however, small insertions and deletions are included in the calculation, identity is reduced to 90.76%. Considering regions that are difficult to align, such as lineage-specific interspersed repeat elements, would further decrease the level of computed identity. Moreover, evolutionary distances exhibit local fluctuations, as in other mammals (3), and less divergence was observed in chromosome X (94.26% identity of aligned bases). The GC-content of the rhesus in aligned bases was not notably lower than that of the human (40.71% versus 40.74%).

Gene content. A human-centric approach was used to generate new macaque gene sets (table S3.1 and fig. S3.1). These sets include (i) Ensembl (12) gene models based primarily on the alignment of the human Uniprot and RefSeq resources with the current assembly to define the overall gene model, followed by the introduction of the macaque-specific sequences (mainly as lineage-specific paralogs) in that framework; (ii) Gnomen (NCBI) models that include the consideration of the available (~50,000) macaque ESTs along with the human RefSeq; and (iii) Nscan data that include multiple-species alignments along with cDNA alignments (13). Overall, ~20,000 loci were predicted by our methods in which at least one exon was found by two additional predictors. An additional ~5000 loci were each predicted by a single method, but manual inspection of a subset of these loci shows that they are enriched in gene-prediction errors, mainly due to misclassification of evidence (e.g., cDNAs from untranslated regions that were classified as containing protein coding). On average, high-confidence orthologs have 97.5% identity between the human and macaque at both the nu-

Fig. 2. Assembly by three methods of the rhesus macaque genome. WGS, whole-genome shotgun. BCM-HGSC, Baylor College of Medicine Human Genome Sequencing Center; WashU-GSC, Washington University Genome Sequencing Center; JCVI, J. Craig Venter Institute. QA/QC, quality assurance and quality control.
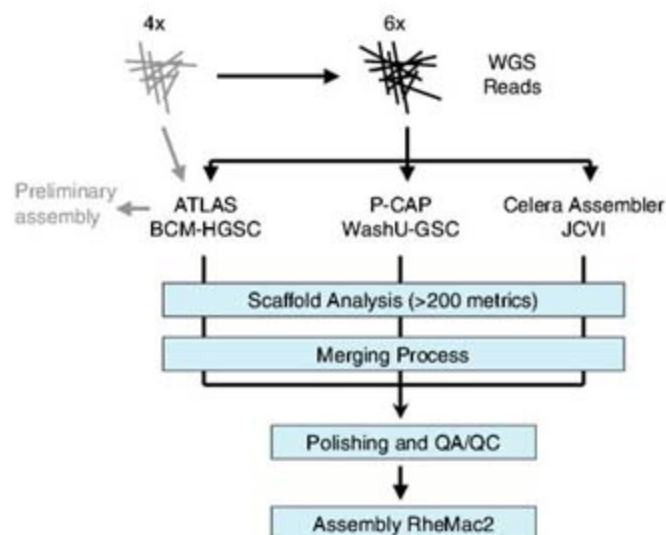
Table 1. *M. mulatta* assembly statistics. Total bases, excluding gaps, number 2,871,189,834.

	Contigs	Scaffolds
Total number	301,039	122,580
N50 size in bp	25,707	24,345,431
Number to N50	32,114	36
Largest in bp	219,335	98,200,701

The Rhesus Macaque Genome

cleotide and amino acid sequence levels. (The nucleotide and amino acid percentages agree because roughly one-third of nucleotide differences within coding regions change an amino acid.)

Overall repetitive landscape. Repeat elements account for ~50% of the genomes of all sequenced primates (*14*) (Table 2). Similar to the human, the rhesus macaque contains about 320,000 recognizable copies from more than 100 different families of DNA transposons and more than half a million recognizable copies of endogenous retroviruses (ERVs). In general, the DNA transposons show no new lineages, but the ERVs demonstrate a complex phylogeny and many examples of new and expanded family members, some resulting from horizontal transmission. In addition, we conservatively estimate that ~20,000 L1s [a family of long interspersed elements (LINEs)], and ~110,000 *Alu* elements [a primate-specific family of short interspersed elements (SINEs)], were specifically acquired in the Old World monkey lineage. These two retrotransposon families accounted for most lineage-specific insertions and have played a major role in shaping genomic architecture. Among them, rhesus macaque–specific subsets (derived from the L1PA5 lineage and *Alu*Y) are frequently polymorphic and can be assayed by polymerase chain re-action (PCR) genotyping analyses for genetic studies (*15*).

Determining Ancestral Genome Structure

Cytogenetically visible rearrangements. The most notable genomic differences among the HCR trio are the presence of cytogenetically visible rearrangements. The human and chimpanzee karyotypes are distinguishable by one chromosome fusion and nine cytogenetically visible pericentric inversions (*16*); with the use of the macaque as an outgroup, all of these breakpoints (except those induced by two inversions) have now been characterized at the DNA sequence level (*17*). Analysis of genomic sequence confirms that 14 breakpoints, corresponding to seven inversions, occurred in the chimpanzee lineage, as indicated in Fig. 1. (Five of the inversions are summarized in table S4.1.) The pericentric inversions of human chromosomes 1 and 18 and the fusion creating human chromosome 2 are specific to the human. Comparison of the reconstructed human-chimpanzee ancestral genome and the rhesus genome reveals 43 breakpoints on the microscopic scale (Figs. 1 and 3).

Submicroscopic rearrangements. Previous analyses [reviewed in (*14*)] have indicated that primate genomes harbor more structural differences than visible by cytogenetic staining. Analysis of these events is complicated by two issues: the draft state of the genomes and the presence of extensive segmental duplications. We analyzed these structural rearrangements by using the distance between orthologous blocks in each species to infer the ancestral genome structure and determine where rearrangements occurred on the phylogenetic tree. We excluded events smaller than 10 kilobase pairs (kbp), which are mostly due to retroposon insertions, and focused on cytogenetically undetectable breakpoints induced by insertions, deletions, inversions, and complex rearrangements of sizes between 10 kbp and 4 Mbp. Data were combined from inversion detection and ancestral reconstructions by the contiguous ancestral regions method (*18*) and gap detection by the genomic triangulation method (*19*), which further integrates data from genomic sequence comparisons (*20*) and comparative maps (*8, 9, 21*). The analysis revealed more than 1000 rearrangement-induced breakpoints through the HCR lineages, of which 820 occur between rhesus and the reconstructed human-chimpanzee ancestor (Fig. 3 and fig. S4.1). Each chromosome therefore constitutes a complex mosaic, with multiple changes introduced to orthologous counterparts. When rhesus macaque is compared with the human-chimpanzee ancestor, the X chromosome exhibits three times more rearrangements per megabase than the autosomes. This is both statistically significant and consistent with a slightly more than threefold difference observed in the human lineage following the branching off of chimpanzee (*19*). Given that a slower rate of variability at the single-nucleotide level in the X chromosome compared with autosomes has been interpreted as support for speciation models, this difference is worthy of further investigation (*22*).

Duplications in the Genome and Gene Family Expansions

Genomic Duplications. Segmental duplication of genomic regions and the genes they contain

Table 2. Summary of repeat content of the rhesus macaque genome compared with the human and chimpanzee genomes. hg18, human genome version 18; panTro2, *Pan troglodytes* version 2; rheMac2, rhesus macaque version 2; LTR, long terminal repeat; MIR, mammalian interspersed repeat. SVA is a composite repetitive element named after its main components, SINE, variable number of tandem repeats, and *Alu*; includes SVA precursor elements.

Species	DNA	LTR/ERV	LINE		SINE		SVA
			L1	L2	*Alu*	MIR	
hg18	355,000	506,000	572,000	363,000	1,144,000	584,000	3400
panTro2	305,000	453,000	558,000	315,000	1,111,000	553,000	4400
rheMac2	327,000	432,000	531,000	298,000	1,094,000	539,000	150

Fig. 3. Chromosomal breakpoints between rhesus macaque and the human-chimpanzee ancestor. Each chromosome is represented by a white bar (left) and a colored bar (right). A total of 820 thin horizontal lines in the white bars represent submicroscopic breakpoints (10-kbp to 4-Mbp range) detected by genomic triangulation (*19*), and 43 thick black lines in the colored bars represent breakpoints on a microscopic scale (>4 Mbp) (*7*). Numbers above each bar show the total lines within the bar.
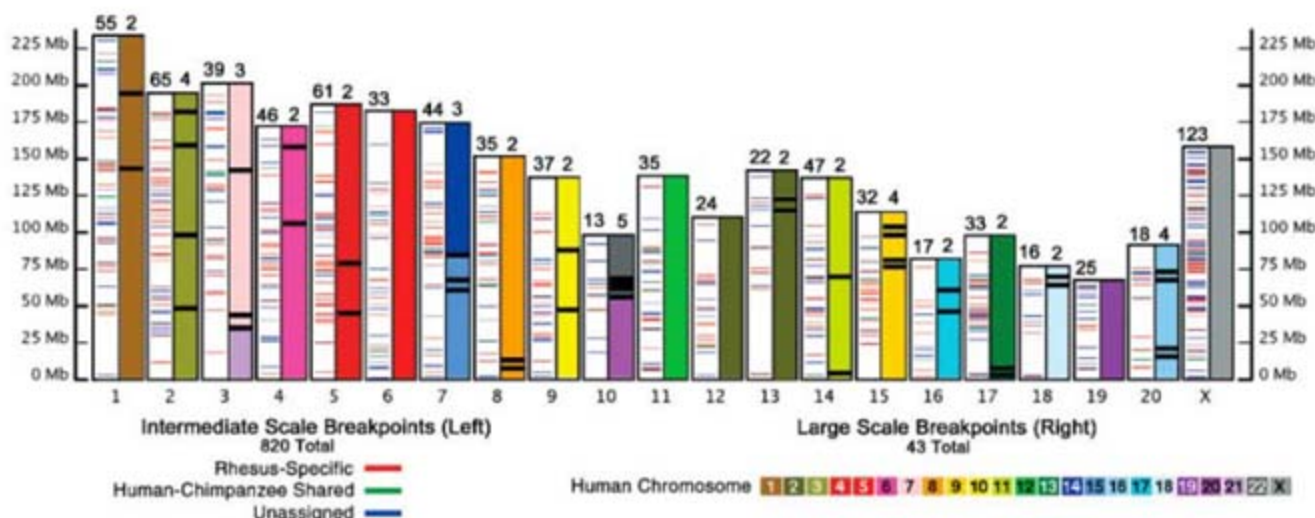
are well known in mammals and are postulated to drive fundamental processes, including the birth of new genes and the subsequent expansion of gene families (23). To discover duplications in the macaque genome, we used a battery of different complementary approaches. Two of these, whole-genome assembly comparison (24) and BLASTZ (25) analysis of segmental duplications, depended directly on the assembly. We used a third method, whole-genome shotgun sequence detection (26), that calculated depth of coverage of the raw shotgun sequence reads relative to the assembly. A fourth procedure was created on the basis of BAC end sequence reads combined with BACs that were directly mapped by means of the pooled genomic indexing method (21). The common interspersed repeat families were not considered in any of these analyses.

The first two approaches identified approximately 35.0 Mb of a recently duplicated sequence in the macaque assembly. A further ~15 Mb were collapsed in the assembly and discovered by whole-genome shotgun sequence detection (fig. S5.1 and table S5.1). Adjusting for these collapsed duplications and the overall assembly coverage, we estimate that approximately 66.7 Mb or 2.3% of the macaque genome consists of segmental duplication (Fig. 4)—this proportion is substantially lower than that of either the human or chimpanzee genome (5 to 6%) (26, 27).

The pooled genomic indexing and BAC end sequence read methods suggested slightly high-er levels of overall duplication, on the basis of fluorescence in situ hybridization analysis of randomly selected large-insert BAC clones (28). However, this estimate was still less than the 4.8% recently estimated for the baboon genome (28). Overall, we consider 2.3% to be the lower bound of duplicated genomic DNA in the macaque genome.

As with the human and chimpanzee, the analysis of the macaque assembly revealed an enrichment of segmental duplications near gaps, centromeres, and telomeres (14, 29). The study also identified segmental duplications that contain genes of high biological significance. For example, the *CCL3L1-CCL4* gene region [for which copy-number variation in humans is correlated with susceptibility to HIV infection (30)], cytochrome P450 (associated with toxicity response), *KRAB-C2H2* zinc finger (a developmental regulatory transcription factor), olfactory receptor (smell), human leukocyte antigen (HLA), and other immune and autoantigen gene families were all observed in regions of genome duplication.

Expansion of gene families. Two approaches were used to study gene family structure directly within the draft genome sequence: (i) a statistical approach, based on a likelihood model of gene gain and loss across the mammalian tree (31) and (ii) hybridization of whole genomic DNA to cDNA arrays [a variation of array-based comparative genomic hybridization (array CGH)] to observe changes in gene content directly (32). The results are shown in Tables 3 and 4.

The statistical approach revealed that 1358 genes were gained by duplication along the macaque lineage. This method simultaneously estimates rates of change along individual lineages and generates a quantitative assessment of confidence in rate differences among lineages. Iterative modeling revealed higher rates in primates, relative to other mammals. The rates are similar to those obtained by independent methods in both humans (33) and rodents (3).

We identified 108 gene families, computationally predicted to have changed in size among the primates, evolving at a significantly higher rate than the overall primate rates of gene gain and loss (all $P < 0.0001$, Table 3). More than 60% of the macaque-specific expansions display evidence of positive selection in their coding sequences, supporting the notion that this rate disparity may be driven by natural selection.

Gene copy-number estimates by genomic hybridization (cDNA array CGH) (32) identified 51 genes (124 cDNAs) with copy-number increases in the macaque, relative to the human (Table 4 and table S5.2). Of these array CGH-predicted macaque-specific increases, 33% (17 out of 51) were also found by computational analysis of gene family gains and losses. A separate analysis found that 55% (28 out of 51) are increased in copy number as estimated by BLAST-like Alignment Tool (BLAT)–based (34) predictions from the rheMac2 assembly. In contrast, when random sets of genes (cDNAs) were chosen for BLAT queries, only 1.45% suggest copy-number increases ($P < 0.0001$).

The genome-wide acceleration identified in primates may be due to an explosion in the number of *Alu* transposable elements in the primate ancestor, which may have allowed an increase in the rates of nonallelic homologous recombination, leading to higher rates of both duplication and deletion (35). Alternatively, the rates of duplicate gene fixation may be due to the small population size in primates (36) relative to rodents.

Particular expanded gene families. Expansion of individual gene families may help to identify processes that distinguish biological features among organisms. One example in humans is the preferentially expressed antigen of melanoma (*PRAME*) gene family that consists of a single gene on chromosome 22q11.22 and a cluster of several dozen genes on chromosome 1p36.21. *PRAME* and PRAME-like genes are actively expressed in cancers but normally manifest testis-specific expression and may thus have a role in spermatogenesis. The genomic organization is complicated; the cluster on human chromosome 1 exhibits copy-number variation in human populations (37, 38) and, together with a similar orthologous cluster on mouse chromosome 4, apparently arose by translocation not long before the divergence of primates and

RheMac2 Assembly
WGAC All (> 1kb > 90%)
34.97 Mb non redundant
Intra: 24.73 Mb nr (blue)
Inter: 13.74 Mb nr (red)

Fig. 4. Global pattern of macaque segmental duplications. The statistics are based on all WGAC duplications (> 90%, >1 kb in length), whereas the figure displays only those between 90 and 95% sequence identity and >10 kb in length for simplicity. Red lines indicate interchromosomal (Inter) duplications, blue ticks show intrachromosomal (Intra) events, and purple bars show centromeric, acrocentric, and/or large-gap regions. WGAC, whole-genome assembly comparison. nr, nonredundant.

The Rhesus Macaque Genome

rodents, about 85 Mya (*39*) (Fig. 5 and fig S5.2). After that translocation event, the human and mouse gene clusters expanded independently. Evidence for positive selection has been found in these genes, and two segmental duplications postdating human-chimpanzee divergence added about a dozen genes to the human cluster.

To properly resolve evolutionary changes in the PRAME gene family, we further sequenced six macaque BAC clones to achieve a higher data quality, and we assembled them into a single contig (table S2.6). These eight PRAME genes were compared with human and chimpanzee genes identified from the latest assemblies for both species. We estimated a phylogeny for all identified genes, designating the mouse gene cluster and the human *PRAME* gene on chromosome 22 as outgroups. We then reconciled this gene tree with the species tree by maximum parsimony. Our reconstruction reveals extensive duplication early in primate evolution (Fig. 5B, branch *a*), in recent chimpanzee evolution (Fig. 5B, branch *d*), and, most notably, in recent human evolution (Fig. 5B, branch *e*). The PRAME gene cluster appears to have been much less dynamic on the macaque lineage (Fig. 5B, branch *b*) and in early hominins (the human and chimpanzee branch, Fig. 5B, branch *c*). A large inverted tandem duplication occurred on the macaque lineage shortly after divergence from the human lineage, but no additional large-scale rearrangements are evident. The relative quiescence in macaque allows us to identify older duplications that are difficult to discern in the exceedingly complex human self-alignments (*7*).

The inferred PRAME gene tree shows pronounced differences in evolutionary rates across branches, as well as some quite long branches that suggest bursts of adaptive change. Using maximum likelihood methods, we found evidence of positive selection on several of these branches (Fig. 5A). This positive selection, combined with the highly variable pattern of gene duplication and expansion, suggests that the PRAME gene family has played a key role in species evolution.

We identified a second segment of extensive genomic duplications concentrated at the telomere of macaque chromosome 9, orthologous to a human locus at 10p15.3 and observed by multiple approaches to be distributed throughout the macaque genome. The genes phosphofructokinase-platelet form (*PFKP*) and *DIP2C* were expanded in this region and yielded the highest array CGH macaque-to-human ratios in the genome (average log$_2$ ratios of 3.30 and 2.54, respectively). *DIP2C* is implicated in segmentation patterning, although its relevance to macaque evolution is currently obscure. *PFKP* is important in sugar (fructose) metabolism, raising the possibility that the pronounced copy-number expansion in macaque may be relevant to the

Table 3. Gene families with significant copy-number expansions ($P < 0.0001$) in the human and the identical statistic for the rhesus macaque. Gene family ID, identification numbers from Ensembl version 41. Family size, number of gene copies in the current genome assemblies. Gains and losses, number of genes gained and lost since the human's split with chimpanzee or the macaque's split with human-chimpanzee lineage. IG, immunoglobulin; IGE, immunoglobulin E; Pre, precursor; MHC, major histocompatibility complex; TCR, T cell receptor; ENV, envelope; ATP, adenosine 5′-triphosphate.

Gene family ID	Description	Family size	Gains	Losses
Expanded in human				
ENSF00000000020	IG heavy chain V region	42	10	0
ENSF00000000073	Receptor	56	16	0
ENSF00000000233	Peptidyl prolyl cis trans isomerase	38	9	0
ENSF00000000312	Histone H2b	28	7	0
ENSF00000000597	Golgin subfamily A	49	26	0
ENSF00000000664	Ankyrin repeat domain	33	9	0
ENSF00000000822	Unknown	15	9	0
ENSF00000000841	Tripartite motif	21	7	1
ENSF00000000936	Centaurin gamma	15	9	0
ENSF00000001036	Cold inducible RNA binding	22	8	0
ENSF00000001546	Ubiquitin carboxyl terminal hydrolase	16	13	2
ENSF00000001599	Leucine-rich repeat	14	7	0
ENSF00000001665	DNA mismatch repair PMS2	12	5	0
ENSF00000001738	Unknown	15	7	0
ENSF00000001920	40S ribosomal S26	13	7	1
ENSF00000001974	Unknown	17	3	0
ENSF00000002160	Double homeobox	15	13	0
ENSF00000002570	Keratin associated 5	7	2	0
ENSF00000003683	Unknown	5	3	0
ENSF00000004835	Ambiguous	13	9	0
Expanded in macaque				
ENSF00000000014	HLA class I	17	12	0
ENSF00000000037	HLA class I	16	10	0
ENSF00000000070	Keratin type I	65	30	0
ENSF00000000077	Histone H3	32	11	0
ENSF00000000085	IG kappa chain V region	47	22	2
ENSF00000000138	Keratin type II	39	10	0
ENSF00000000150	Taste receptor type 2	23	9	0
ENSF00000000178	Aldo keto reductase family 1	19	9	0
ENSF00000000397	Ral guanine nucleotide dissoc stim.	19	10	1
ENSF00000000432	Killer cell IG receptor Pre MHC class I	9	3	0
ENSF00000000630	TCR beta chain V region Pre	18	9	0
ENSF00000000705	ENV polyprotein	13	11	0
ENSF00000000766	60S ribosomal l7A	26	17	1
ENSF00000000773	Ribosomal l7	23	12	0
ENSF00000000826	60S ribosomal l23A	20	6	0
ENSF00000001027	60S ribosomal l17	12	3	0
ENSF00000001077	Nucleoplasmin	17	9	0
ENSF00000001211	67-kD laminin	18	10	0
ENSF00000001235	Nonhistone chromosomal HMG 17	24	12	0
ENSF00000001236	60S ribosomal l31	23	11	0
ENSF00000001249	60S ribosomal l12	16	8	0
ENSF00000001359	USP6 N terminal	14	10	0
ENSF00000001460	Prohibitin	7	4	0
ENSF00000001671	60S ribosomal l32	10	6	0
ENSF00000001861	40S ribosomal S10	9	5	0
ENSF00000002239	60S ribosomal l19	8	5	0
ENSF00000002279	40S ribosomal S17	8	4	0
ENSF00000002476	60S ribosomal l18	7	4	0
ENSF00000002633	IGE binding	19	14	0
ENSF00000003321	Argininosuccinate synthase	9	6	0
ENSF00000003395	10-kD heat shock protein	11	8	0
ENSF00000004083	ATP synthase subunit G	4	3	0
ENSF00000007347	Unknown	7	3	0

Table 4. Genes identified as expanded in copy number in the macaque, relative to the human, by the array CGH method. The leftmost column represents IMAGE cDNA clones that show array CGH–predicted copy number increases in the rhesus macaque relative to the human. The middle two columns list corresponding gene names and array CGH $\log_2$ macaque-to-human ratios. The rightmost column presents BLAT-predicted copy numbers based on rheMac2 and hg18 genome assemblies.

IMAGE clone	Gene	Average $\log_2$ array CGH ratio	rheMac2/hg18 BLAT-predicted copy numbers
41109/1900937	PFKP	3.30	3/2†
454926/1862434	DIP2C	2.54	4/2†
1475421/757369	EST	1.74	7/4†
50877/110020	EST	1.48	3/4
795258/1574131/191978	ATP5J2	1.42	29/10†
824545/278888	EST	1.37	0/1
2457916/322067	DNAJC8	1.37	9/6†
504421/435036	ADFP	1.27	6/4†
769921/146882	UBE2C	1.17	8/3†
155620/154809	IGL	1.14	8/10
1985794*/1470105	EST	1.14	1/6
32083/270786	FLJ30436	1.13	2/3
306344/773260	MAT2B	1.11	3/2†
884480/194908	COX7C/PRO2463	1.09	13/4†
244205*/462961/768172/824776*/123971	DHFR	1.05	14/13†
72745/1626871*	HLA	1.02	1/5
1493107/1637726	LTB4DH/EST	0.97	3/2†
163407/843374	STOM	0.96	2/2
258666/428043	PSMB7/EST	0.96	3/2†
112498/824894	EST	0.95	0/0
32231/770984	EST/FLJ12442	0.95	3/2†
981713*/953542*/981925*	EST	0.95	0/0
1636233/814459	C9orf23	0.93	4/2†
529185/609265	SELK	0.93	6/4†
298965/1472754/512003*	COX6B	0.93	7/4†
322561/240620*	EST	0.92	31/22†
208656/415195	FLJ20294	0.92	2/2
840698/39977	FLJ20254/MAPRE3	0.90	4/2†
773287/1635681	NDUFA2	0.89	4/2†
756763*/725401*	EST	0.85	1/1
80742/80694	EST	0.85	0/0
1415672*/1558664*	EST	0.84	0/0
323806/38029	EST	0.83	2/2
595547/997889*	EST	0.83	1/1
953654*/953643*	EST	0.83	0/0
783035*/783249*	EST	0.83	1/1
884272*/1415750*	H3F3A	0.83	45/40†
322175/210873	EST/PPY2	0.81	1/1
1569731/1569604	EST	0.79	4/4
292982/129431	EST	0.78	1/2
112785/361565*	RoXaN/GLUD1	0.78	7/4†
292452/450327	SMBP	0.78	2/2
1606275/1534633	C6orf129/STOM	0.77	3/2†
212847*/1415750*	EST	0.76	22/16†
664121*/745347*	PIG7/EST	0.76	4/2†
982122/982113/121546/503715	EST/FLJ14668	0.73	9/6†
950688/811603	EST/ATP6V1G1	0.73	4/3†
327202/194384	EST/BTF3	0.72	1/1
897007/897676*	EST	0.71	1/1
301388/825470	TOP2A	0.69	6/2†
590390/756469*	RoXaN	0.62	3/2†

*Consistent with computational analysis of gene family gains and losses.   †BLAT-based copy-number estimates of rheMac2 and hg18 genome assemblies that are consistent with array CGH predictions.

high-fruit diet common among macaques. As with other array CGH copy-number estimates, the functional status of the additional copies is not known. Six of the individual macaque BACs that mapped to the region revealed related duplicated sequences on rhesus chromosome 3, which formed from the fusion of orthologs of human chromosomes 7 and 21, suggesting that these genes may have played a role in this expansion.

Another macaque-specific increase involves the 22 HLA-related genes located in the region orthologous to human chromosome 6p21 (table S5.4). A previous study found that HLA gene copy number was higher in the macaque than in the human (40), and our results confirm and extend this finding, demonstrating that the macaque HLA copy number is greater than that found for the human as well as all four great ape species (fig S5.3). This finding also suggests that, although the macaque has been extensively used to model the human immune response, there may be substantial and previously unappreciated differences in HLA function between these species. Notably, the copy number of another immune system–related gene cluster, immunoglobulin lambda-like (IGL) at 22q11.23, is also predicted to be increased in the macaque (table S5.4). Members of the IGL locus encode light chain subunits that are part of the Pre–B cell receptor; do not undergo rearrangements; and, when mutated, can result in B cell deficiency and agammaglobulinemia. Additional known genes predicted by array CGH to have markedly increased copy numbers in the macaque relative to the human include DHFR, ATP5J2, DNAJC8, ADFP, and MAT2B. Overall, the main characteristics of the set of amplified genes were their diversity and the wide variety of genomic regions they occupied.

Orthologous Relationships

The macaque genome has also allowed for a detailed study of more subtle changes that have accumulated within orthologous primate genes. The average human gene differs from its ortholog in the macaque by 12 nonsynonymous and 22 synonymous substitutions, whereas it differs from its ortholog in the chimpanzee by fewer than three nonsynonymous and five synonymous substitutions. Similarly, 89% of human-macaque orthologs differ at the amino acid level, as compared with only 71% of human-chimpanzee orthologs. Thus, the chimpanzee and human genomes are in many ways too similar for characterizing protein-coding evolution in primates, but the added divergence of the macaque helps substantially in clarifying the signatures of natural selection.

General characteristics of orthologous genes. We developed an automatic pipeline to identify 10,376 trios of HCR genes to which we could assign a high confidence of 1:1:1 orthology. For comparison, we also identified 6762 hu-

The Rhesus Macaque Genome
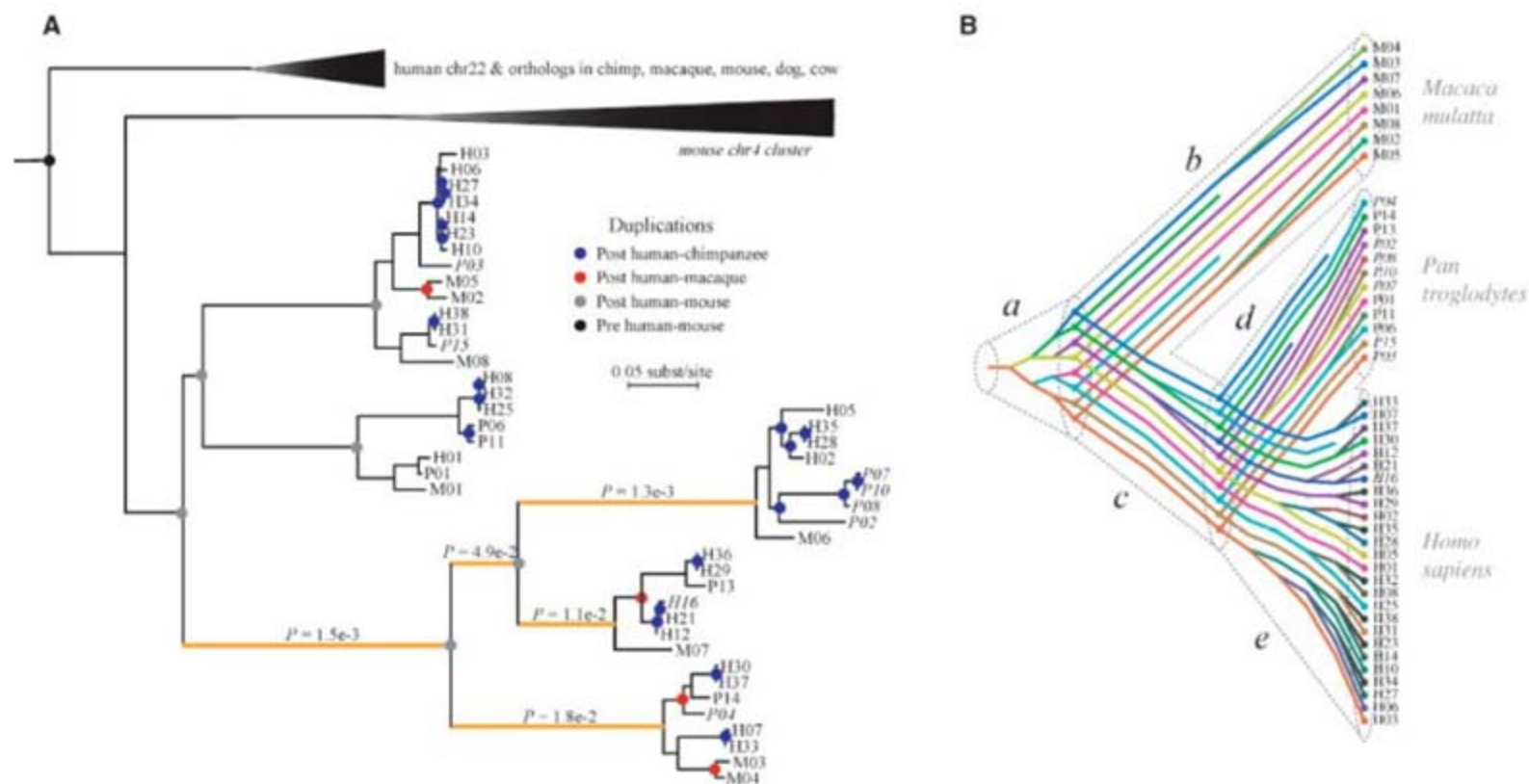
Fig. 5. Organization of the PRAME gene cluster in the HCR lineages. (**A**) Maximum-likelihood phylogeny for PRAME-like genes in the human (H), chimpanzee (P), and rhesus macaque (M) genomes. Colored circles indicate inferred duplication events, partial genes are shown in italics, and branches showing significant evidence of positive selection are colored orange (*P* values are shown above orange lines). Scale bar, 0.05 substitutions per site. (**B**) Another view of the same phylogeny, showing the duplication history in the context of the species tree (7).

man, macaque, mouse, and rat quartets; 5641 HCR, mouse, and rat quintets; and 5286 HCR, mouse, and dog quintets. Because the human gene models are by far the best characterized for primates, we first identified a set of 21,256 known human protein-coding genes derived from a union of the RefSeq (*41*), Vega (*42*), and University of California–Santa Cruz Known Genes (*43*) collections. These genes were then mapped to synteny-based genome-wide multiple alignments (*44, 45*) and subjected to a series of rigorous filters to eliminate spurious annotations, paralogous alignments, genes that have become pseudogenized in one or more species, and genes with incompletely conserved exon-intron structures (*7*). The genes that pass all filters represent 1:1:1 orthologs in which aligned protein-coding bases are highly likely to encode proteins in all species, with identical reading frames.

Despite the draft quality of the chimpanzee and macaque assemblies, the majority of human genes mapped through syntenic alignments to the chimpanzee (93% of genes) and macaque (89%) genomes (Fig. 6) (*7*), and most of these genes were completely alignable in their coding regions. Fairly large fractions of human genes, however, were discarded because of apparent frame-shift insertions and deletions (indels) or nonconserved exon-intron structures with respect to their putative chimpanzee or macaque orthologs. On the basis of 81 finished BACS covering 294 genes, we estimate that, out of

5526 genes failing the filters for alignment completeness, frame-shift indels, and conserved exon-intron structure, 2138 (39%) were discarded completely because of flaws in the macaque assembly; the remaining 3388 (61%) were discarded either because of genuine changes to genes or because of annotation or alignment errors (*7*). Another 2261 genes passed the human-macaque filters but failed the human-chimpanzee filters, and a large majority of these failures were probably due to flaws in the chimpanzee assembly. Altogether, we estimate that finished genomes for the macaque and chimpanzee would allow the number of genes in high-confidence orthologous trios to be increased by at least 23%, to ~12,800 (*7*). Notably, our conservative ortholog sets may create a bias against fast-evolving genes and therefore may lead to underestimates of average levels of divergence and the prevalence of positive selection.

Alignments of the 10,376 orthologous trios were used to estimate the ratio of the rates of nonsynonymous and synonymous substitutions per gene (denoted ω), with continuous-time Markov models of codon evolution and maximum likelihood methods for parameter estimation (*46–48*). This yielded a mean estimate of ω = 0.247 (median 0.144), close to the value of 0.23 estimated for human and chimpanzee genes (*29*). About 9.8% of all genes show no nonsynonymous changes in the three

species, and 2.8% have ω > 1, suggesting that they are under positive selection. Consistent with previous studies (*49*), certain classes of genes exhibit unusually large or small ω values, such as those assigned to the gene ontology (*50*) category "immune response," which have an ω distribution shifted significantly toward larger values, and those assigned to the "transcription factor activity" category, which have a distribution shifted toward smaller values (fig. S6.1).

Our estimates for ω in primates are considerably larger than previously reported estimates for rodents, which have a median of 0.11 (*3*), and larger than similar estimates from primate-versus-rodent comparisons (*29*) (Fig. 7). To compare the average rates of evolution of protein-coding genes in primates with those in other mammals, we estimated a separate value of ω for each branch of a five-species phylogeny, pooling data from all 5286 one-to-one orthologs for these species (fig. S6.2). We obtained similar estimates of ω for the human (ω = 0.169) and chimpanzee (ω = 0.175) lineages, but substantially smaller estimates for the branches leading to nonprimate mammals (ω = 0.104 to 0.128), suggesting a reduction in purifying selection in hominins (*29*). The estimate of ω for the macaque lineage (ω = 0.124) is substantially smaller than the estimates for the human and chimpanzee and is closer to the estimates for the mouse and dog, perhaps reflecting the larger population size of

Fig. 6. Numbers of human genes passing successive filters in the orthology analysis pipeline. Genes are required to fall in regions of large-scale synteny between genomes, to have completely aligned coding regions, not to have frame-shift indels or altered gene structures, and not to show signs of recent duplication.

Legend:
- filter 1: lack of synteny
- filter 2: incomplete alignments
- filter 3: frame-shift indels
- filter 4: changes in exon-intron structure
- filter 5: recent duplication
- remaining 1:1 orthologs

X-axis categories: human chimp; human macaque; human mouse; human rat; human dog; human chimp macaque; human macaque mouse rat; human chimp macaque mouse rat; human chimp macaque mouse dog

macaques compared with the other primates. The estimates for the internal branches between the most recent common ancestors of the human and mouse and of the human and macaque, as well as the most recent common ancestors of the human and macaque and of the human and chimpanzee, are nearly equal to the macaque estimate. This suggests that protein-coding sequence evolution in macaques may have occurred at a typical primate rate, whereas it is the elevated rates in hominins that may be anomalous.

When primate and rodent ω of individual genes were compared, primate orthologs were found to be evolving more rapidly by a 3:2 ratio. This asymmetry was also evident among genes showing substantial differences in primate ω (ω_p), on the basis of human-macaque alignments, and rodent ω (ω_r), deduced from mouse-rat alignments. According to a strict Bonferroni correction for multiple testing, 22 genes showed statistically significant $\omega_p > \omega_r$, whereas only three genes showed $\omega_r > \omega_p$ (McNemar $P < 0.001$). If multiple testing criteria are relaxed, the bias toward larger ω_p is more notable (144 versus 8; tables S6.1 and S6.2). Cases of $\omega_p > \omega_r$ generally reflect an increase in ω_p, whereas cases of $\omega_r > \omega_p$ result both from an increase in ω_r and a decrease in ω_p. The genes showing statistically significant $\omega_p > \omega_r$ are enriched for functions in sensory perception of smell and taste as well as for regulation of transcription (7).

Positive selection. Taking advantage of the additional phylogenetic information provided by the macaque genome, we performed a genome-wide scan for positive selection, using our 10,376 HCR orthologous trios and likelihood ratio tests (LRTs) (51–53). Four different LRTs were performed: test TA, for positive selection across all branches of the phylogeny, and tests TH, TC, and TM for positive selection on the individual branches to human,

Fig. 7. Distributions of ω in primates versus rodents. Histogram of estimates of $\omega = dN/dS$ for human, chimpanzee, and macaque versus estimates for mouse and rat in 5641 orthologous quintets, showing a pronounced shift toward larger values in primates ($P = 2.2 \times 10^{-16}$, Mann Whitney test). Genes with $dN = 0$ or $dS = 0$ are counted in the relative frequencies but not shown.
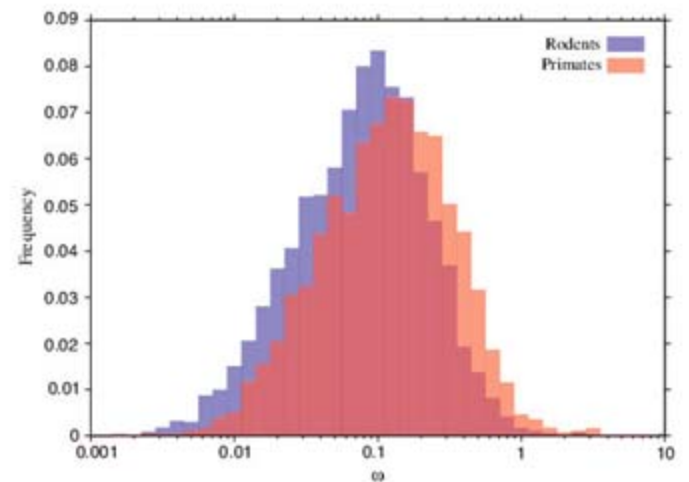
chimpanzee, and macaque, respectively. Our methods use an unrooted tree and cannot distinguish between the branches to macaque and the human-chimpanzee ancestor; for convenience, we refer to the combined branch as the macaque branch. In all cases, variation among sites in ω was allowed and, to reduce the number of parameters to estimate per gene, the branch-length proportions and transition-transversion ratio (κ) were estimated by pooling data from genes of similar G+C content (7). Test TA identified 67 genes, and tests TH, TC, and TM identified 2, 14, and 131 genes (false-discovery rate (FDR) < 0.1 in all cases), respectively. The large number of genes identified for the macaque branch is partly a reflection of its greater length compared with the chimpanzee and human branches (7).

These four sets of genes overlap considerably, particularly among their highest scoring predictions (Table 5 and table S6.3). Their union contains 178 genes, or 1.7% of all genes tested. The two genes identified by TH—those encoding the leukocyte immunoglobulin-like receptor LILRB1 and hypothetical protein LOC399947—were also identified by TA,

and the gene for LILRB1 was identified by TC as well, indicating evidence of positive selection on multiple branches. However, 12 out of 14 genes identified by TC were not identified by the other tests, indicating possible lineage-specific selection in the chimpanzee. These include sex comb on midleg-like 1 (*SCML1*) and protamine 1 (*PRM1*), which were previously identified in an analysis that could not distinguish between selection on the human and chimpanzee branches (52). In addition, 99 genes were identified by TM but not the other tests. These genes may be under lineage-specific selection in the macaque and/or may have experienced positive selection on the branch leading to the most recent common ancestor of the human and chimpanzee.

The genes identified by our tests for positive selection are enriched for several categories from the gene ontology (50) and Protein Analysis Through Evolutionary Relationships (PANTHER) (54) classification systems that are similar to those observed in previous genome-wide scans for positive selection (52, 53). These include defense response, immune response, T cell-mediated immunity, signal transduction,

The Rhesus Macaque Genome

and cell adhesion (tables S6.4 to S6.7). Among the genes in these categories are several immunoglobulin-like genes, including those that encode the leukocyte-associated inhibitory receptors LILRB1 and LAIR1 (located in a cluster on chromosome 19), the T cell surface glycoprotein CD3 epsilon chain precursor CD3E, and the intercellular adhesion molecule 1 precursor ICAM1. Other identified genes associated with cell adhesion and/or signal transduction include those that encode DSG1, a calcium-binding transmembrane component of desmosomes, and the transmembrane protein TSPAN8 (which has gained an exon by duplication in the macaque genome). Genes encoding membrane proteins in general are strongly overrepresented; other examples include the genes that encode connexin 40.1, active in cell communication, and *OPN1SW*, the gene encoding blue-sensitive opsin.

In addition, we observed strong enrichments for new categories such as iron ion binding [e.g., the beta globin (*HBB*), lactotransferrin (*LTF*), and cytochrome B-245 heavy chain genes (*CYBB*)] and oxidoreductase activity (e.g., *KRTAP*5-8 and *KRTAP*5-4, which encode keratin-associated proteins, and *NDUFS*5, which encodes a subunit of the nicotinamide adenine dinucleotide ubiquinone oxidoreductase). Two keratin genes, which are important for hair-shaft formation,

are present among the top-scoring genes; these genes could conceivably have come under positive selection as a result of mate selection or climate change. Genes classified as part of the extracellular region, which include the keratin genes, are in general overrepresented. Many of the identified genes from this category encode secreted proteins, such as the interferon alpha 8 precursor IFNA8, which exhibits antiviral activity; the interleukin 8 precursor IL8, a mediator of inflammatory response; and CRISP1, which is expressed in the epididymis and plays a role at fertilization in sperm-egg fusion.

We found only weak enrichments for genes involved in apoptosis and spermatogenesis (*52*), but we did see a significant excess of high likelihood ratios among genes involved in fertilization. Other categories that show an excess of high likelihood ratios but that are not enriched for genes identified by our tests include blood coagulation, response to wounding, and related categories; epidermis morphogenesis; KRAB-box transcription factor; and olfactory receptor activity (tables S6.6 and S6.7). Their elevated likelihood ratios may reflect either weak positive selection or relaxation of constraint.

The inclusion of the macaque genome substantially improves statistical power to detect positive selection in primates, compared with

previous scans that used only the human and chimpanzee genomes (*29*, *52*). By examining about 8000 human-chimpanzee alignments with a similar LRT, Nielsen *et al.* (*52*) were able to identify only 35 genes with nominal $P < 0.05$, and when considering multiple comparisons, they were able to establish only that a 5% false discovery rate set was nonempty. By contrast, the use of the macaque genome allows the identification of 15 genes under positive selection in hominins and an additional 163 under selection on one or more other branches of the phylogeny, with FDR < 0.1. We estimate that including the macaque genome makes test TA about three times as powerful. However, including macaque rather than mouse (*53*) as an outgroup improves the power of test TH only marginally (*7*).

The genes identified by the LRTs are generally randomly distributed in the genome, and no significant clustering was observed when tested ($P = 0.24$), although small clusters were found on human chromosomes 11 and 19 (*7*). Chromosome 11, with 10 genes identified by test TA, has more than twice the expected number of genes under positive selection, but this enrichment is not significant after correcting for multiple comparisons [$P = 0.10$, Fisher's exact test and Holm correction (*7*)]. However, a significant enrichment was observed for genes overlapping segmental duplications that oc-

Table 5. Selected genes from top 40 showing evidence of positive selection in primates. Accession, the number of the reference transcript for each gene (human). Chr, human chromosome on which reference gene resides. *P* value, nominal *P* value for test TA (*7*). Genes shown have FDR < 0.04. Test, the test (other than test TA) that detected the given gene. The Dup column has a checkmark if a gene overlaps a segmental duplication preceding the human/macaque divergence.

Accession	Gene name	Chr	Description	P value	Test	Dup
AB126077	*KRTAP5-8*	11	Keratin-associated protein 5-8	6.20×10^{-16}	TM	✓
NM_006669	*LILRB1*	19	Leukocyte immunoglobulin-like receptor	7.20×10^{-14}	TH, TC	✓
NM_001942	*DSG1*	18	Desmoglein 1 preproprotein	1.10×10^{-10}	–	
NM_173523	*MAGEB6*	X	Melanoma antigen family B, 6	5.30×10^{-8}	TC	✓
NM_054032	*MRGPRX4*	11	G protein–coupled receptor MRGX4	5.60×10^{-8}	TM	✓
NM_000397	*CYBB*	X	Cytochrome b-245, beta polypeptide	1.50×10^{-7}	TM	
NM_001911	*CTSG*	14	Cathepsin G preproprotein	1.50×10^{-7}	TM	
NM_000735	*CGA*	6	Glycoprotein hormones, alpha polypeptide	1.20×10^{-6}	TM	
NM_001012709	*KRTAP5-4*	11	Keratin-associated protein 5-4	2.70×10^{-6}	TM	✓
NM_000201	*ICAM1*	19	Intercellular adhesion molecule 1 precursor	2.70×10^{-6}	TM	
NM_001131	*CRISP1*	6	Acidic epididymal glycoprotein-like 1 isoform 1	1.60×10^{-5}	TM	
NM_002287	*LAIR1*	19	Leukocyte-associated immunoglobulin-like	3.10×10^{-5}	TM	✓
NM_153368	*CX40.1*	10	Connexin40.1	4.90×10^{-5}	–	
NM_018643	*TREM1*	6	Triggering receptor expressed on myeloid cells	6.30×10^{-5}	TM	
NM_000300	*PLA2G2A*	1	Phospholipase A2, group IIA	1.30×10^{-4}	–	
BC020840	*TCRA*	14	T cell receptor alpha chain C region	1.50×10^{-4}	–	
NM_000733	*CD3E*	11	CD3E antigen, epsilon polypeptide	1.50×10^{-4}	TM	
NM_001014975	*CFH*	1	Complement factor H isoform b precursor	1.50×10^{-4}	–	
NM_001423	*EMP1*	12	Epithelial membrane protein 1	1.50×10^{-4}	TM	
NM_001424	*EMP2*	16	Epithelial membrane protein 2	1.50×10^{-4}	TM	
NM_002170	*IFNA8*	9	Interferon, alpha 8	1.50×10^{-4}	–	
NM_030766	*BCL2L14*	12	BCL2-like 14 isoform 2	1.50×10^{-4}	–	
NM_006464	*TGOLN2*	2	Trans-golgi network protein 2	1.80×10^{-4}	TM	
NM_014317	*PDSS1*	10	Prenyl diphosphate synthase, subunit 1	1.80×10^{-4}	–	
NM_000518	*HBB*	11	Beta globin	2.00×10^{-4}	TM	

curred before the human-macaque divergence ($P = 0.006$, Fisher's exact test), suggesting an increased likelihood of adaptive evolution following gene duplication. Four of the top five genes identified by test TA overlap segmental duplications that predate the human-macaque divergence (Table 5).

Genetic Variation in Macaques

The use of rhesus macaques as animal models of human physiology can be greatly enhanced by an improved understanding of their underlying genetic variation. To explore rhesus genetic diversity and to create resources for further genetic studies, we generated a total of 26.2 Mb of whole-genome shotgun sequence from 16 unrelated individuals (eight of Chinese origin and eight of Indian origin, table S7.1). We next identified 26,479 single-base differences [putative single-nucleotide polymorphisms (SNPs)] through comparison with the reference genome. Overall, we found approximately one SNP per kilobase, which is on average close to that found in similar human studies. There was a surprising difference of 50% in overall diversity between the autosomes and the X chromosome (Fig. 8A); we expected a value of 75%. This expectation was based on differences in effective chromosome population sizes, given that females have two X chromosomes and males carry only one. The reduction in diversity could be due to recent selective sweeps of positively selected recessive mutations on the X chromosome (*55*).

We also found that the frequency of the whole-genome shotgun SNPs differed substantially among the animals from the different populations (0.95/kb in Indian rhesus and 1.06/kb in Chinese rhesus), and there was suggestive variation in SNP density within their subpopulations (SD = 0.0275/kb for Chinese macaques; SD = 0.0527/kb for Indian macaques). Together with complementary data from PCR analysis of polymorphic L1 and *Alu* element insertions (figs. S7.1 and S7.2) that showed population sub-

structure, this prompted additional experiments in which 48 animals from the two populations were surveyed by PCR-direct DNA sequencing. Details and most conclusions from that study have been reported by Hernandez *et al.* (*56*), including a demonstration that >67% of SNPs discovered by direct sequencing are private to each subpopulation. The strong population differentiation is reflected in fixation index (FST) values (a measure of population differentiation) and a marked difference in Watterson's (*57*) estimate of the population mutation rate between the two groups. Here, we observed that the population differences are also reflected in differential distribution of Tajima's *D* statistic and in linkage disequilibrium across sampled regions (Fig. 8, B and C). Each of these statistics further reflects the possibilities of sweeps of natural selection or major differences in population histories that must be factored into ongoing genetic studies. These initial insights into the underlying patterns of variation within individual animals will therefore provide the basis for future genetic analyses. In addition to their utility for identification of individual animals, the SNP markers will be invaluable for larger-scale population studies.

Male mutation bias. A comparison of human-rhesus substitution rates (calculated at interspersed repetitive elements) between the X chromosome and the autosomes yielded an estimate of the male-to-female mutation rate ratio (α) of 2.87 (95% CI = 2.37 to 3.81; table S7.2). This value is lower than α = 6 estimated for the human and chimpanzee (*58*) but higher than α = 2 estimated for the mouse and rat (*3, 59*). Thus, this argues against a uniform magnitude of male mutation bias in mammals (*5*) and supports a correlation between male mutation bias and generation time (*60, 61*).

Human Disease Orthologs in the Macaque

While the general morphological and physiological similarities between humans and macaques greatly enhance the utility of the latter as

a model organism, specific differences in their underlying coding sequences can also provide biological insights. By comparing human disease genes with their macaque equivalents, we identified numerous instances in which the allele observed in the macaque corresponds to the disease allele in the human. These occurrences suggest that the human disease variants could be either persistent (i.e., ancestral) or recurring sequences that represent the recapitulation of ancestral states that may once have been protective, but which now result in adverse consequences for human health (*62*).

To identify the ancestral disease-associated alleles in human, we screened the macaque and chimpanzee assemblies for the presence of any of the 64,251 different disease-causing or disease-associated mutations collected in the Human Gene Mutation Database (*63, 64*). A total of 229 substitutions were identified for which the amino acid considered to be mutant in human corresponded to the wild-type amino acid present in macaque, chimpanzee, and/or a reconstructed ancestral genome (Table 6) (*65*) (see table S8.1 for a full list).

One surprising result of the analysis was the identification of several human loci that, when mutated, give rise to profound clinical phenotypes, including severe mental retardation. For example, the macaque data revealed deleterious alleles in the ornithine transcarbamylase (OTC) and phenylalanine hydroxylase (PAH) genes, which are associated in human with OTC deficiency and phenylketonuria. In humans, these mutations greatly perturb the normal serum amino acid levels. Direct examination of macaque blood revealed lower concentrations of cystine and cysteine than in the human and slightly higher concentrations of glycine than in the human, but no increase in phenylalanine or ammonia, which might have been a predicted result of these changes (tables S8.2 and S8.3). Although the effect of the observed alleles might be greatly influenced by compensatory mutations (*66*) or other environ-
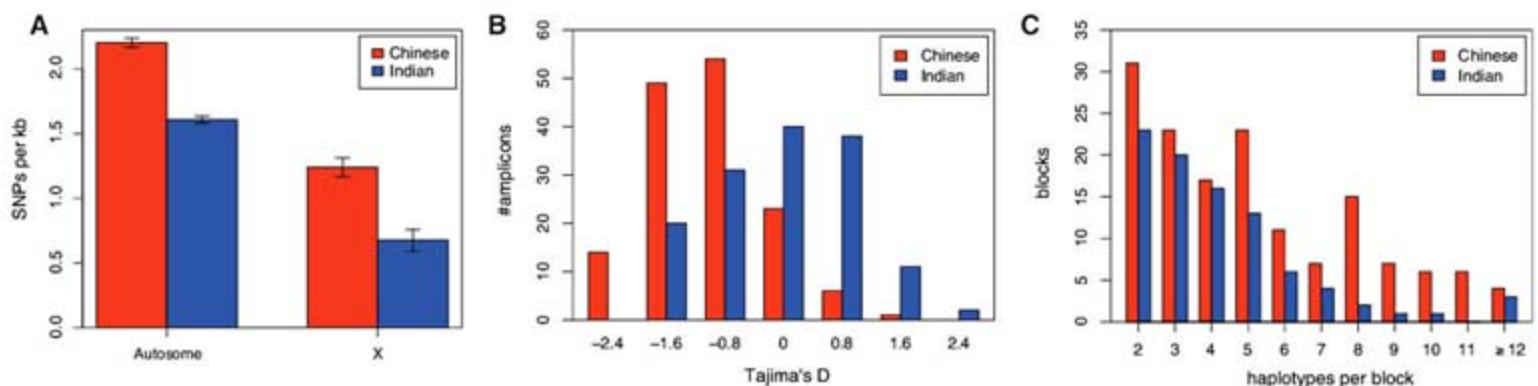
Fig. 8. SNP within rhesus macaques. (**A**) SNP densities per kilobase for eight Chinese (blue) and eight Indian (red) individuals in autosomes and the X chromosome. Error bars indicate standard error with variance calculated across individual-chromosome replicates. (**B**) Distribution of Tajima's *D* statistic across 166 amplicons for each population ($n = 38$ for Indian and $n = 9$ for Chinese individuals). (**C**) The distribution of the number of haplotypes per haplotype block (determined using the four-gamete test) across five regions.

mental factors, it remains a possibility that the basic metabolic machinery of the macaque may exhibit functionally important differences with respect to our own (Fig. 9).

Ancestral mutations were also identified in the N-alpha-acetylglucosaminidase (NAGLU) gene that gives rise to mucopolysaccharidoses (Sanfillipo syndrome), which is also characterized by profound mental retardation. Their occurrence invites further investigation of the contribution of this and related genes to the phenotypic differences between macaques and humans, and the potential for further exploration of these monkeys as models for this disorder.

We also identified a human mutation associated with Stargardt disease and macular dystrophy that matches an ancestral allele by replacing lysine with glutamine at position 223 of the human ABCA4 protein (Fig. 9). Umeda et al. (67) reported the presence of the glutamine in a cynomolgus monkey, and all other eutherian mammals as well as the predicted boreoeutherian ancestral sequence have glutamine at this position. Furthermore, glutamine is present at this residue in Xenopus, thereby implying conservation through some 300 million years of vertebrate evolution. Thus, it may be inferred that the ancestral glutamine has been replaced by lysine in humans. Similarly, one CFTR mutation [Phe^{87}→Leu^{87} (Phe87Leu)] is present not only in most mammals (Fig. 9) but also in Fugu, also implying extensive conservation through vertebrate evolution.

Impact of a Genomic Sequence on Biological Studies

In addition to its impact on comparative and genetic studies, the genome sequence reported here heralds a new era in laboratory studies of macaque biology. The full potential for more precise definition of this animal model and its gene content is not yet realized, but the value of the new sequence in guiding DNA microarrays for studying macaque gene expression has already become clear (68). Previously, human or macaque EST-based arrays had been used for expression studies (69). The most recently released microarray now adds probes designed by alignment of the 3′ untranslated regions of 23,000 human RefSeq genes to the sequences from the initial macaque genome release (January 2005, Mmul0.1, approximate genome coverage of 3.5-fold). The vast majority of the probes on this array (98.5%) now match the current macaque genome release with high confidence and represent 18,690 unique genomic loci. These provide a representation of recognized functional pathways with an enhancement about three times that of the previous data, and overall more uniform and robust hybridization signals compared with those of previous microarrays (69) (tables S9.1 to S9.3).

The power of global transcriptional profiling with advanced macaque-specific reagents has been demonstrated in studies of virulence and pathogenicity of influenza from historic pandemic strains, as well as from emerging agents of zoonotic origin. We infected macaques with the human influenza strain A/Texas/36/9 (70) and compared the expression changes observed in lung tissues to those seen in whole blood during the course of infection. Figure 10 shows a differential time course of expression between interferon-induced genes and genes in the inflammation pathway, in different tissues (table S9.4). The increased expression in lung tissue shortly after infection reflects the early innate response, whereas genes associated with the reemergence of the inflammation pattern at day 7 implicate a transition to an adaptive immune response. These kinds of studies will be crucial for elucidating all of the transitions from innate to adaptive immune responses and are fully enabled by the macaque-specific microarrays developed from the genome sequences.

We expect many more immediate examples of the impact of other tools developed from the finished macaque genome. For example, the requirement for improvements in PCR-based methods is shown by a recent report on the large-scale cloning of terminal exons for macaque genes, in which the use of human primers was successful, on average, in 67% of cases (71). Only a native sequence can allow sufficient precision for these types of highly specific assays. A similar increase of activity in studies of the macaque proteome can be predicted, given that early efforts in macaque proteomics have had to rely on human reference sequences for analyzing liquid chromatography and tandem mass spectrometry data (70).

Discussion

The draft genomic sequence reported here has already moved the macaque from a model that has been much studied at the level of physiology, behavior, and ecology to a whole-organism system that can be interrogated at the level of the single DNA base. This transformation is evident in the literature as well as in this special section (15, 19, 57, 72).

Additional general conclusions emerged from this study. First, the data make it conceivable to define completely all of the operational components of the pathways underlying the individual biological systems that together constitute the functioning adult macaque. For example, a complete description of all the different macaque immune function components will enable an even more thoughtful use of rhesus macaques in areas such as AIDS research and for vaccine production.

Second, we were struck by the high value of adding regions of genome finishing to the draft sequence for the comparative analyses of genes and duplicated structures. This provides an argument for future finished primate genomes.

Third, the data now provide new opportunities to explore the basic biology of this highly successful species. Rhesus macaques retain a broad geographic distribution with reasonably healthy population numbers and widely studied ecology and ethology. The genetic resources generated in this study will

Table 6. Examples of human mutations that cause inherited disease and match an ancestral or nonhuman primate state. Chr:start-stop shows the address in the March 2006 human assembly. Name is the name used by the Human Gene Mutation Database (64). The notation "N>A:CHMT" means that N is the consensus human amino acid, A is the disease-associated form, C is in the current chimp assembly, H is in the inferred human-chimp ancestor, M is in rhesus, and T is in the inferred human-rhesus ancestor (the mouse and dog were used as outgroup species) (73).

Chr:start-stop	Strand	Name	Replacement N>A:CHMT	Gene	Disease
chr1:94270150-94270152	-	CM014300	R>Q:RRQR	ABCA4	Stargardt disease
chr1:94316821-94316823	-	CM015072	H>R:RRRR	ABCA4	Stargardt disease
chr1:94337037-94337039	-	CM042258	K>Q:QQQQ	ABCA4	Stargardt disease
chr6:26201158-26201160	+	HM030028	V>A:VVAA	HFE	Hemochromatosis
chr7:116936418-116936420	+	CM940237	F>L:FFLL	CFTR	Cystic fibrosis
chr7:117054872-117054874	+	CM941984	K>R:KKRK	CFTR	Cystic fibrosis
chr12:101761685-101761687	-	CM962547	Y>H:YYHY	PAH	Phenylketonuria
chr12:101784521-101784523	-	CM941128	I>T:IITI	PAH	Phenylketonuria
chr13:51413354-51413356	-	CM044579	V>A:AAAA	ATP7B	Wilson disease
chr13:112843266-112843268	+	CM021094	D>E:DDED	F10	Factor X deficiency
chr17:37948991-37948993	+	CM040465	R>Q:RRQQ	NAGLU	Sanfilippo syndrome B
chr19:43656115-43656117	+	CM064230	S>G:GGGG	RYR1	Malignant hyperthermia
chrX:38111528-38111530	+	CM941115	R>H:RRHH	OTC	Ornithine hyperammonemia
chrX:38125613-38125615	+	CM961052	T>M:MTTT	OTC	Ornithine hyperammonemia
chrX:138458220-138458222	+	CM045148	E>K:EEKK	F9	Hemophilia B

undoubtedly form the basis of many analyses of population variability and inter-population diversity.

Finally, the genomic rearrangements, duplications, gene-specific expansions, and measurements of the impact of natural selection presented

A Ancestral alleles are now mutations in human

	ABCA4 Lys223Gln	CFTR Phe87Leu
boreoeutherian	R V T Q A G R	Y G I L L Y L
primate	. . . .	. . . .
catarrhine	. . . .	. . . .
hominid	. . . .	. . . F . . .
human mutation	. . . .	. . . .
human normal	. . . K . . .	. . . F . . .
chimp	. . . .	. . . F . . .
macaque	. . . .	. . . .
rat	. . A . . . W	. . V . . . .
mouse	. . . .	. . . .
rodent	. . . . . . W	. . . .
cow	. . . . . . A .	. . . I . . .
dog	. . . .	. . . .
elephant	. . . .	. . . .
opossum	. . G H L . M	H . . I . . .

B Altered sequence is disease-associated in human but normal in macaque

	OTC Arg40His
boreoeutherian	L K G R D L L
primate	. . . .
catarrhine	. . . .
hominid	. . . .
human mutation	. . . H . . .
human normal	. . . .
chimp	. . . .
macaque	. . . H . . .
rat	. . . .
mouse	. . . .
rodent	. . . .
cow	. . . .
dog	. . . .
elephant	. . . .
opossum	M . . . H . .

Fig. 9. Ancestral disease mutations. Examples of human mutations that match the sequences of chimp and/or macaque are shown. (**A**) Genes in which the ancestral allele is now the disease-associated allele in humans. (**B**) An instance in which the mutant allele in humans is the normal allele in macaque. The amino acid sequences predicted for the boreoeutherian ancestor (65) are given on the top row of each alignment block. Identities are shown as dots and differences are given as letters (73). The position of the mutation in humans is boxed in orange, and the box extends through the relevant comparisons.
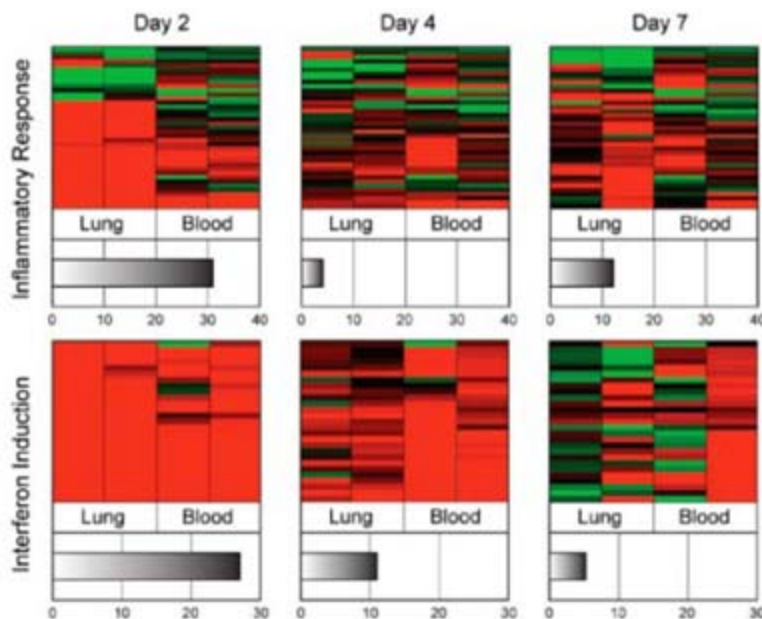
here have revealed the rich and heterogeneous genomic changes that have occurred during the evolution of the human, chimpanzee, and macaque. The marked diversity of the types of change that have occurred demonstrate a major feature of primate evolution: The aggregation of changes that we see, even in closely related species, does not reflect smooth, progressive, and orderly genomic divergence. Models of abrupt or punctuated evolution already acknowledge that smooth and continuous change is difficult to achieve on an evolutionary time scale, but this study provides a notable example of the operation of this principle in our close relatives.

References and Notes

1. C. Groves, *Primate Taxonomy* (Smithsonian Institution Press, Washington, DC, 2001).
2. S. Kumar, S. B. Hedges, *Nature* **392**, 917 (1998).
3. R. A. Gibbs et al., *Nature* **428**, 493 (2004).
4. F. C. Chen, W. H. Li, *Am. J. Hum. Genet.* **68**, 444 (2001).
5. N. Patterson, D. J. Richter, S. Gnerre, E. S. Lander, D. Reich, *Nature* **441**, 1103 (2006).
6. L. M. Zahn, B. R. Jasny, Eds., poster from the special issue on the Macaque Genome, *Science* **316**, following p. 246 (13 April 2007); interactive online (www.sciencemag.org/sciext/macaqueposter/).
7. Materials, methods, and additional discussion are available on *Science* Online.
8. J. Rogers et al., *Genomics* **87**, 30 (2006).
9. W. J. Murphy et al., *Genomics* **86**, 383 (2005).
10. B. Dutrillaux, *Hum. Genet.* **48**, 251 (1979).
11. J. Wienberg, R. Stanyon, A. Jauch, T. Cremer, *Chromosoma* **101**, 265 (1992).
12. T. J. Hubbard et al., *Nucleic Acids Res.* **35**, D610 (2007).
13. M. J. van Baren, M. R. Brent, *Genome Res.* **16**, 678 (2006).
14. International Human Genome Sequencing Consortium, *Nature* **409**, 860 (2001).
15. K. Han et al., *Science* **316**, 238 (2007).
16. J. J. Yunis, O. Prakash, *Science* **215**, 1525 (1982).
17. H. Kehrer-Sawatzki, D. N. Cooper, *Hum. Genet.* **120**, 759 (2007).
18. J. Ma et al., *Genome Res.* **16**, 1557 (2006).
19. R. A. Harris et al., *Science* **316**, 235 (2007).
20. K. J. Kalafus, A. R. Jackson, A. Milosavljevic, *Genome Res.* **14**, 672 (2004).
21. A. Milosavljevic et al., *Genome Res.* **15**, 292 (2005).
22. N. H. Barton, *Curr. Biol.* **16**, R647 (2006).
23. J. A. Bailey, E. E. Eichler, *Nat. Rev. Genet.* **7**, 552 (2006).
24. J. A. Bailey, A. M. Yavor, H. F. Massa, B. J. Trask, E. E. Eichler, *Genome Res.* **11**, 1005 (2001).
25. S. Schwartz et al., *Genome Res.* **13**, 103 (2003).
26. J. A. Bailey et al., *Science* **297**, 1003 (2002).
27. Z. Cheng et al., *Nature* **437**, 88 (2005).
28. X. She et al., *Genome Res.* **16**, 576 (2006).
29. The Chimpanzee Sequencing and Analysis Consortium, *Nature* **437**, 69 (2005).
30. E. Gonzalez et al., *Science* **307**, 1434 (2005).
31. M. W. Hahn, B. T. De, J. E. Stajich, C. Nguyen, N. Cristianini, *Genome Res.* **15**, 1153 (2005).
32. A. Fortna et al., *PLoS Biol.* **2**, E207 (2004).
33. M. Lynch, J. S. Conery, *J. Struct. Funct. Genomics* **3**, 35 (2003).
34. W. J. Kent, *Genome Res.* **12**, 656 (2002).
35. J. A. Bailey, E. E. Eichler, *Cold Spring Harb. Symp. Quant. Biol.* **68**, 115 (2003).
36. M. Lynch, M. O'Hely, B. Walsh, A. Force, *Genetics* **159**, 1789 (2001).
37. A. J. Iafrate et al., *Nat. Genet.* **36**, 949 (2004).

Fig. 10. Application of rhesus-specific microarrays. A microarray based on the rhesus macaque draft genome was used to analyze gene expression in a macaque model of human influenza infection. Gray bars measure an overall response for indicated functional categories, based on corresponding heat maps, and reveal a significant rebound in expression at day 7 for genes associated with the inflammatory response, when compared to interferon induction. Red, increased expression; green, reduced expression. Details are given in (7, 70).

38. J. Sebat et al., Science 305, 525 (2004).
39. Z. Birtle, L. Goodstadt, C. Ponting, BMC Genomics 6, 120 (2005).
40. R. Daza-Vamenta, G. Glusman, L. Rowen, B. Guthrie, D. E. Geraghty, Genome Res. 14, 1501 (2004).
41. K. D. Pruitt, T. Tatusova, D. R. Maglott, Nucleic Acids Res. 35, D61 (2007).
42. J. L. Ashurst et al., Nucleic Acids Res. 33, D459 (2005).
43. F. Hsu et al., Bioinformatics 22, 1036 (2006).
44. M. Blanchette et al., Genome Res. 14, 708 (2004).
45. W. J. Kent, R. Baertsch, A. Hinrichs, W. Miller, D. Haussler, Proc. Natl. Acad. Sci. U.S.A 100, 11484 (2003).
46. N. Goldman, Z. Yang, Mol. Biol. Evol. 11, 725 (1994).
47. R. Nielsen, Z. Yang, Genetics 148, 929 (1998).
48. Z. Yang, Comput. Appl. Biosci. 13, 555 (1997).
49. D. Graur, L. Li, Fundamentals of Molecular Evolution (Sinauer Associates, Tel Aviv, Israel, ed. 2, 2000).
50. M. Ashburner et al., Nat. Genet. 25, 25 (2000).
51. Z. Yang, R. Nielsen, Mol. Biol. Evol. 19, 908 (2002).
52. R. Nielsen et al., PLoS Biol. 3, e170 (2005).
53. A. G. Clark et al., Science 302, 1960 (2003).
54. H. Mi et al., Nucleic Acids Res. 33, D284 (2005).
55. A. J. Betancourt, Y. Kim, H. A. Orr, Genetics 168, 2261 (2004).
56. R. D. Hernandez et al., Science 316, XXXX (2007).
57. G. A. Watterson, Theor. Popul. Biol. 7, 256 (1975).
58. J. Taylor, S. Tyekucheva, M. Zody, F. Chiaromonte, K. D. Makova, Mol. Biol. Evol. 23, 565 (2006).
59. K. D. Makova, S. Yang, F. Chiaromonte, Genome Res. 14, 567 (2004).
60. A. Bartosch-Harlid, S. Berlin, N. G. Smith, A. P. Moller, H. Ellegren, Evolution Int. J. Org. Evolution 57, 2398 (2003).
61. M. P. Goetting-Minesky, K. D. Makova, J. Mol. Evol. 63, 537 (2006).
62. J. V. Neel, Am. J. Hum. Genet. 14, 353 (1962).
63. P. D. Stenson et al., Hum. Mutat. 21, 577 (2003).
64. Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff (www.hgmd.org; October 2006 release).
65. M. Blanchette, E. D. Green, W. Miller, D. Haussler, Genome Res. 14, 2412 (2004).
66. L. Gao, J. Zhang, Trends Genet. 19, 678 (2003).
67. S. Umeda et al., Invest. Ophthalmol Vis. Sci. 46, 683 (2005).
68. J. D. Chismar et al., Biotechniques 33, 516 (2002).
69. J. C. Wallace et al., BMC Genomics 8, 28 (2007).
70. T. Baas et al., J. Virol. 80, 10813 (2006).
71. E. R. Spindel et al., BMC Genomics 6, 160 (2005).
72. M. Ventura et al., Science 316, 243 (2007).
73. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
74. This project was supported by National Human Genome Research Institute grants to the Baylor College of Medicine Human Genome Sequencing Center (U54 HG003273), Washington University Genome Sequencing Center (U54 HG003079), and the J. Craig Venter Institute (U54 HG003068). We thank members of the NHGRI staff for their ongoing efforts: A. Felsenfeld, J. Peterson, M. Guyer, and W. Lu. Additional acknowledgments of support are available online (7). We thank the California National Primate Research Center (NPRC), Oregon NPRC, Southwest NPRC, and Yerkes NPRC for contributing biological samples used in this study.

Rhesus Macaque Genome Sequencing and Analysis Consortium

Project Leader: Richard A. Gibbs[1,2]

White paper: Jeffrey Rogers,[3] Michael G. Katze,[4] Roger Bumgarner,[4] Richard A. Gibbs,[1,2] George M. Weinstock[1,2]

Principal investigators: Richard A. Gibbs,[1,2] Elaine R. Mardis,[5] Karin A. Remington,[6] Robert L. Strausberg,[6] J. Craig Venter,[6] George M. Weinstock,[1,2] Richard K. Wilson[5]

Analysis leaders: Mark A. Batzer,[7] Carlos D. Bustamante,[8] Evan E. Eichler,[9] Richard A. Gibbs,[1,2] Matthew W. Hahn,[10] Ross C. Hardison,[11] Kateryna D. Makova,[11] Webb Miller,[11] Aleksandar Milosavljevic,[1,2] Robert E. Palermo,[4] Adam Siepel,[8] James M. Sikela,[12] George M. Weinstock[1,2]

Genome sequencing: Tony Attaway,[1,2] Stephanie Bell,[1,2] Kelly E. Bernard,[5] Christian J. Buhay,[1,2] Mimi N. Chandrabose,[1,2] Marvin Dao,[1,2] Clay Davis,[1,2] Kimberly D. Delehaunty,[5] Yan Ding,[1,2] Huyen H. Dinh,[1,2] Shannon Dugan-Rocha,[1,2] Lucinda A. Fulton,[5] Ramatu Ayiesha Gabisi,[1,2] Toni T. Gamer,[1,2] Richard A. Gibbs,[1,2] Jennifer Godfrey,[5] Alicia C. Hawes,[1,2] Judith Hernandez,[1,2] Sandra Hines,[1,2] Michael Holder,[1,2] Jennifer Hume,[1,2] Shalini N. Jhangiani,[1,2] Vandita Joshi,[1,2] Ziad Mohid Khan,[1,2] Ewen F. Kirkness[6] (leader), Andrew Cree,[1,2] R. Gerald Fowler,[1,2] Sandra Lee,[1,2] Lora R. Lewis,[1,2] Zhangwan Li,[1,2] Yih-shin Liu,[1,2] Stephanie M. Moore,[1,2] Donna Muzny[1,2] (leader), Lynne V. Nazareth[1,2] (leader), Dinh Ngoc Ngo,[1,2] Geoffrey O. Okwuonu,[1,2] Grace Pai,[5] David Parker,[1,2] Heidie A. Paul,[1,2] Cynthia Pfannkoch,[6] Craig S. Pohl,[5] Yu-Hui Rogers,[6] San Juana Ruiz,[1,2] Aniko Sabo,[1,2] Jireh Santibanez,[1,2] Brian W. Schneider,[1,2] Scott M. Smith,[5] Erica Sodergren,[1,2] Amanda F. Svatek,[1,2] Teresa R. Utterback,[1,2] Selina Vattathil,[1,2] Wesley Warren[5] (leader), George M. Weinstock,[1,2] Courtney Sherell White[1,2]

Genome assembly: Asif T. Chinwalla[5] (leader), Yucheng Feng,[5] Aaron L. Halpem,[6] LaDeana W. Hillier,[5] Xiaoqiu Huang,[13] Ewen F. Kirkness,[6] Pat Minx,[5] Joanne O. Nelson,[5] Kymberlie H. Pepin,[5] Xiang Qin,[1,2] Karin A. Remington,[6] Granger G. Sutton[6] (leader), Eli Venter,[6] Brian P. Walenz,[6] John W. Wallis,[5] George M. Weinstock,[1,2] Kim C. Worley[1,2] (leader), Shiaw-Pyng Yang[5]

Mapping: LaDeana W. Hillier,[5] Steven M. Jones,[14] Marco A. Marra,[14] Mariano Rocchi,[15] Jacqueline E. Schein,[14] John W. Wallis[5]

Sequence finishing: Christian J. Buhay,[1,2] Yan Ding,[1,2] Shannon Dugan-Rocha,[1,2] Alicia C. Hawes,[1,2] Judith Hernandez,[1,2] Michael Holder,[1,2] Jennifer Hume,[1,2] Ziad Mohid Khan,[1,2] Zhangwan Li,[1,2] Dinh Ngoc Ngo,[1,2] Aniko Sabo[1,2]

Assembly comparison: Robert Baertsch,[16] Asif T. Chinwalla,[5] Laura Clarke,[17] Miklós Csürös,[18] Jarret Glasscock,[5] R. Alan Harris,[1,2] Paul Havlak,[1,2] LaDeana W. Hillier,[5] Andrew R. Jackson,[1,2] Huaiyang Jiang,[1,2] Yue Liu,[1,2] David N. Messina,[5] Xiang Qin,[1,2] Yufeng Shen,[1,2] Henry Xing-Zhi Song,[1,2] George M. Weinstock[1,2] (leader), Kim C. Worley[1,2] (leader), Todd Wylie,[5] Lan Zhang[1,2]

Gene prediction: Ewan Birney,[17] Laura Clarke[17]

Repetitive elements: Mark A. Batzer[7] (leader), Kyudong Han,[7] Miriam K. Konkel,[7] Jungnam Lee,[7] Webb Miller,[11] Arian F. A. Smit,[19] Brygg Ullmer,[20] Hui Wang,[7] Jinchuan Xing[7,21]

Ancestral genomes and segmental duplications: Richard Burhans,[11] Ze Cheng,[9] Miklós Csürös,[18] Evan E. Eichler,[9] R. Alan Harris,[1,2] Andrew R. Jackson,[1,2] John E. Karro,[11] Jian Ma,[22] Aleksandar Milosavljevic[1,2] (leader), Brian Raney,[22] Xinwei She[9]

Gene duplication/gene families: Michael J. Cox,[12] Jeffery P. Demuth,[10] Laura J. Dumas,[12] Matthew W. Hahn[10] (leader), Sang-Gook Han,[10] Janet Hopkins,[12] Anis Karimpour-Fard,[23] Young H. Kim,[24] Jonathan R. Pollack,[24] James M. Sikela[12] (leader)

PRAME Gene Family Analysis: Webb Miller[11] (leader), Donna Muzny,[1,2] Brian Raney,[22] Aniko Sabo,[1,2] Adam Siepel,[8] Tomas Vinar[8]

Orthologous genes: Charles Addo-Quaye,[11] Jeremiah Degenhardt,[8] Alexandra Denby,[8] Melissa J. Hubisz,[25] Amit Indap,[8] Carolin Kosiol,[8] Bruce T. Lahn,[25,26] Heather A. Lawson,[11] Alison Marklein,[8] Rasmus Nielsen,[27] Adam Siepel[8] (leader), Eric J. Vallender,[25,26] Tomas Vinar[8]

Population genetics: Mark A. Batzer[7] (leader), Carlos D. Bustamante[8] (leader), Andrew G. Clark,[28] Jeremiah Degenhardt,[8] Betsy Ferguson,[29] Richard A. Gibbs,[1,2] Matthew W. Hahn,[10] Kyudong Han,[7] Ryan D. Hernandez,[8] Kashif Hirani,[1,2] Amit Indap,[8] Hildegard Kehrer-Sawatzki,[30] Jessica Kolb,[8] Miriam K. Konkel,[7] Jungnam Lee,[7] Lynne V. Nazareth,[1,2] Shobha Patil,[1,2] Ling-Ling Pu,[1,2] Jeffrey Rogers,[3] Yanru Ren,[1,2] David Glenn Smith,[3] Brygg Ullmer,[20] Hui Wang,[7] David A. Wheeler,[1,2] Jinchuan Xing[7,21]

Sex chromosome evolution: Kateryna D. Makova,[11] Ian Schenck[11]

Human disease orthologs: Edward V. Ball,[31] Rui Chen,[1,2] David N. Cooper,[31] Belinda Giardine,[11] Richard A. Gibbs,[1,2] Ross C. Hardison[11] (leader), Fan Hsu,[22] W. James Kent,[22] Arthur Lesk,[11] Webb Miller,[11] David L. Nelson,[2] William E. O'Brien,[2] Kay Prüfer,[32] Peter D. Stenson[31]

Additional biological impact of genomic sequence: Michael G. Katze,[4] Robert E. Palermo[4] (leader), James C. Wallace[4]

Macaque sample collection: Hui Ke,[33] Xiao-Ming Liu,[34] Peng Wang,[33] Andy Peng Xiang,[33] Fan Yang[33]

Genome browser: Robert Baertsch,[16] Galt P. Barber,[22] David Haussler[35,16] (leader), Donna Karolchik,[22] Andy D. Kern,[22] Robert M. Kuhn,[22] Kayla E. Smith,[22] Ann S. Zwieg[22]

[1]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA. [2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. [3]Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX 78227, USA. [4]Department of Microbiology, University of Washington, Seattle, WA 98195, USA. [5]Genome Sequencing Center, Washington University, St. Louis, MO 63108, USA. [6]J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA. [7]Department of Biological Sciences, Biological Computation and Visualization Center, Center for BioModular Multi-scale Systems, Louisiana State University, Baton Rouge, LA 70803, USA. [8]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA. [9]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. [10]Department of Biology and School of Informatics, Indiana University, Bloomington, IN 47405, USA. [11]Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, University Park, PA 16802, USA. [12]Human Medical Genetics and Neuroscience Programs, Department of Pharmacology, University of Colorado at Denver and Health Sciences Center, Aurora, CO 80045, USA. [13]Department of Computer Science, Iowa State University, Ames, IA 50011, USA. [14]Genome Sciences Centre, British Columbia Cancer Agency, 570 West 7th Avenue, Vancouver, BC, Canada. [15]Department of Genetics and Microbiology, University of Bari, Bari, Italy. [16]Department of Bioinformatics, University of California Santa Cruz, Santa Cruz, CA 95060, USA. [17]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. [18]Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, Montréal, QC H3C 3J7, Canada. [19]Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103–8904, USA. [20]Center for Computation and Technology, Department of Computer Sciences, Louisiana State University, Baton Rouge, LA 70803, USA. [21]Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA. [22]Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, USA. [23]Department of Preventative Medicine and Biometrics, University of Colorado at Denver and Health Sciences Center, Aurora, CO 80045, USA. [24]Department of Pathology, Stanford University, Stanford, CA 94305, USA. [25]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA. [26]Howard Hughes Medical Institute, Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA. [27]Institute of Biology, University of Copenhagen, Copenhagen DK-1017, Denmark. [28]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA. [29]Genetics Research and Informatics Program, Oregon National Primate Research Center, Beaverton, OR 97006, USA. [30]Institute of Human Genetics, University of Ulm, Ulm, 89081, Germany. [31]Institute of Medical Genetics, Cardiff University, Heath Park, Cardiff, CF14 4XN, UK. [32]Department Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, 04103, Germany. [33]Centre for Stem Cell Biology and Tissue Engineering, Sun Yat-sen University, Guangzhou 510080, China. [34]South-China Primate Research and Development Center, Guangzhou 510080, China. [35]Howard Hughes Medical Institute, Santa Cruz, CA 95060, USA.

REPORT

Human-Specific Changes of Genome Structure Detected by Genomic Triangulation

R. A. Harris,[1,2] J. Rogers,[3] A. Milosavljevic[1,2]*

Knowledge of the rhesus macaque genome sequence enables reconstruction of the ancestral state of the human genome before the divergence of chimpanzees. However, the draft quality of nonhuman primate genome assemblies challenges the ability of current methods to detect insertions, deletions, and copy-number variations between humans, chimpanzees, and rhesus macaques and hinders the identification of evolutionary changes between these species. Because of the abundance of segmental duplications, genome comparisons require the integration of genomic assemblies and data from large-insert clones, linkage maps, and radiation hybrid maps. With genomic triangulation, an integrative method that reconstructs ancestral states and the structural evolution of genomes, we identified 130 human-specific breakpoints in genome structure due to rearrangements at an intermediate scale (10 kilobases to 4 megabases), including 64 insertions affecting 58 genes. Comparison with a human structural polymorphism database indicates that many of the rearrangements are polymorphic.

The human, chimpanzee, and rhesus macaque genomes have now all been sequenced, but the draft quality of the nonhuman primate genomes hampers accurate comparisons between these species. A challenge in using draft-quality genome assemblies to analyze evolutionary changes is that relatively large numbers of differences less than 4 Mb in size among closely related organisms, due to microinsertions, deletions, and copy-number variations, occur in highly duplicated regions that are hard to assemble and compare. Lineage-specific rearrangements in primates have been detected by chromosome banding (*1*) and fluorescence in situ hybridization (*2*), but these methods are not sensitive enough to identify microinsertions and deletions. Array comparative genomic hybridization can detect microinsertions and deletions (*3–6*) but cannot position them on the genome. Chimpanzee fosmid end sequences (FESs) that map inconsistently, based on insert size and orientation, onto the human genome have been used to detect structural differences (*7*), but such studies are prone to false positives and negatives in regions where the genome assembly is incomplete or erroneous and are limited to comparisons of two species. FES mapping, along with pairwise genome assembly comparisons between human and chimpanzee

(*8*), could not establish the ancestral genome state nor assign structural rearrangements to specific lineages. It is increasingly evident that the reconstruction of primate genome evolution requires the integration of genomic data obtained by different methodologies (*9–11*), due, in part, to segmental duplications that confound both genome assembly and determination of orthology. We performed a three-way comparison of the rhesus, human, and chimpanzee genomes to reconstruct the ancestral genome at the branching point of chimpanzee and human. We were able to infer human-specific rearrangements, including insertions, deletions, and inversions, at an intermediate scale of resolution above the 10-kb retroposon insertion size and below the 4-Mb size detectable by cytogenetic methods.

Genomic triangulation reconstructs genome evolution by inputting genomes and other available genomic data from at least three species, two of which form a monophyletic group and one of which is basal to those two and serves as an outgroup. The output is a reconstruction of the ancestral genome of the monophyletic group and a map of breakpoints in specific branches of an unrooted phylogenetic tree. Genomic triangulation consists of three main steps (Fig. 1 and SOM methods): (i) blockset construction, (ii) ancestral threading, and (iii) ancestral gapset construction.

The blockset construction step (Fig. 1) takes two genomes and produces blocks consisting of pairs of orthologous chromosomal segments unbroken by large-scale rearrangements (*12*), each block corresponding to a segment of the ancestral genome of the monophyletic group. Blocks are inferred from collinear orthologous anchors

across two extant genomes. The orthologous anchors are markers in either comparative linkage maps or radiation hybrid maps or are inferred from genome assemblies and large-insert clones. When clone ends of one species anchor onto loci in another species at a distance consistent with insert size and in correct orientation, a clone-sized block is inferred. Overlapping consistent clone-sized blocks are merged into larger blocks. Blocks produced by different anchoring methods are similarly merged on the basis of regions where their genomic coordinates overlap. If different methods produce inconsistent blocks, the longest block is chosen as the best orthology assignment.

The ancestral threading step (Fig. 1) deduces the ancestral genome structure from blocksets in a process similar to standard genome assembly. The overlap of blocks from different blocksets is deduced from overlaps of positional coordinates of the blocks within a single species. All overlapping blocks from the three pairwise blocksets are "threaded" by this method into scaffolds of an ancestral genome. The inferred ancestral genome and the pairwise blocksets between extant species are then used to deduce ancestral blocksets connecting the extant genomes with the ancestral genome.

Because every breakpoint is flanked by two blocks, for each pair of adjacent blocks we define an entity we call a gap. The distances between the paired blocks in the two genomes are associated with each gap. The ancestral gapset construction step (Fig. 1) infers ancestral gapsets from ancestral blocksets. The ancestral gaps localize breakpoints positionally within a genome and evolutionarily within a branch of the phylogenetic tree.

Genomic triangulation was applied to map the evolution of the genomes of the human, chimpanzee, and rhesus macaque. The results, integrated with those obtained by other methods, are summarized in (*13*). The input data for genomic triangulation were human, chimpanzee, and rhesus genome assemblies anchored by Pash (*14*) and University of California Santa Cruz (UCSC) Alignment Nets (*15*), mapped fosmid and bacterial artificial chromosome (BAC) end sequences, rhesus BACs mapped by pooled genomic indexing (*16*), and human-rhesus comparative linkage (*17*) and radiation hybrid (*18*) maps. Varying combinations of merged data sets were input to detect breakpoints at six different levels of resolution (table S4). Detected breakpoints were merged into a single set for further analysis. A total of 288 putative human-specific breakpoints were detected by genomic triangulation. Figure 1 shows the general method that allowed us to infer the existence of a human-specific breakpoint.

Visual inspection of the breakpoint loci by means of the UCSC Genome Browser revealed 158 breakpoints that could not be corroborated

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. [2]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA. [3]Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX 78245, USA.

*To whom correspondence should be addressed.

The Rhesus Macaque Genome

and were classified as possible false positives (SOM methods and table S5). Of these, 31 were due to gaps in the human assembly. The remaining 127 are disagreements with alignment nets in the UCSC Genome Browser and may represent artifacts associated with the current implementation of the genomic triangulation method. Among the remaining 130 human-specific breakpoints, there are 64 insertions, 7 deletions, 16 inversions, and 43 breakpoints that could not be unambiguously assigned to a specific type of rearrangement (Fig. 2). The average size of detected rearrangements on the human genome was 110,063 base pairs (bp), ranging from 20 to 1,365,171 bp.

The human-specific rearrangements were compared with rearrangements detected by the mapping of chimpanzee FESs onto the human genome (7). A total of 52 breakpoints detected by genomic triangulation overlapped with 68 rearrangements, out of a total of 592, detected by FES mapping. Because FES mapping does not assign rearrangements to a specific lineage, by virtue of overlap the 68 rearrangements detected by FES mapping could now be placed in the human lineage. In order to estimate the amount of overlap expected by chance, 100 sets of random genomic locations, with sizes matched to the rearrangements detected by genomic triangulation, were created and their overlap with the FES rearrangements was determined. On average, 25 random locations overlapped with FES rearrangements, indicating a 2.1-fold (52/25) enrichment for overlap in the genomic triangulation set as compared to the random set.

Comparison of detected rearrangements to the Segmental Dups track from the UCSC Genome Browser (19) revealed that 1 deletion, 7 inversions, and 32 unclassified rearrangements were within 10 kb of a segmental duplication. Of the 64 insertions, 42 (66%) overlapped segmental duplications, with 25 insertions being at least 98% identical to their paralogs, as would be expected for recent duplication events (Fig. 2). The remaining 20 insertions were mostly under 20 kb, with 17 consisting of retrotransposons flanking a small segment of putatively transposed sequence, suggesting transduction. The insertions account for roughly one-half of the 8 Mb (8) of human-specific sequence estimated from comparisons of human and chimpanzee genome assemblies, but only a small fraction of the 37 Mb of human-specific sequence estimated on the basis of whole-genome shotgun-read analysis (20).

To further substantiate the insertions, chimpanzee and rhesus BAC and fosmid clone end sequences with mappings that span the insertions were identified. Chimpanzee or rhesus clone ends that map onto the human genome at a distance significantly greater than expected based on the distribution of the clone insert sizes suggest that an insert has occurred in the human

relative to the other species. Of 64 insertions, 22 were spanned by both chimpanzee and rhesus clones and 24 by clones from only one of the species, with a mapping distance significantly longer than expected.

A total of 22 of 130 (17%) breakpoints occur on the X chromosome, which comprises only 5% of the genome. This highly significant enrichment (z score > 6) is consistent with a threefold enrichment in breakpoints in X detected in rhesus-human comparisons (13) and contrasts both with the significantly lower base pair–level divergence of the primate X chromosome as compared to the autosomes (21–23) and with the conservation of the cytogenetically detectable structure of the X chromosome in primates.

We identified 58 genes affected by insertions (table S8), including 36 gene copies fully contained within insertions. An additional 22 genes were either partially duplicated or contained an insertion. Of the 36 fully duplicated genes, 7 were previously identified as human-specific (20).

To determine whether the detected human-specific rearrangements are fixed in all humans, we identified overlaps between the rearrangements and structural variants from The Centre for Applied Genomics (TCAG) Database of Genomic Variants (24). Of the 130 rearrangements, 53 overlapped with at least one polymorphism on the basis of coordinates. We also randomly selected 100 sets of 130 genomic locations, matching the genomic triangulation–detected rearrangement sizes, and identified overlap with the TCAG database. On average, only 28 random locations overlapped with polymorphisms. This 1.9-fold (53/28) enrichment in polymorphisms in the detected set as compared to the random set suggests that a significant fraction of rearrangements are structural alleles that are either polymorphic, and hence not fixed, in all humans or else are sites of recurrent rearrangements.

Genomic triangulation clearly mimics the overlap and layout stages of the overlap–layout–consensus genome assembly method (25), but the two methods differently employ abductive (hypothesis-generating) and deductive inference.
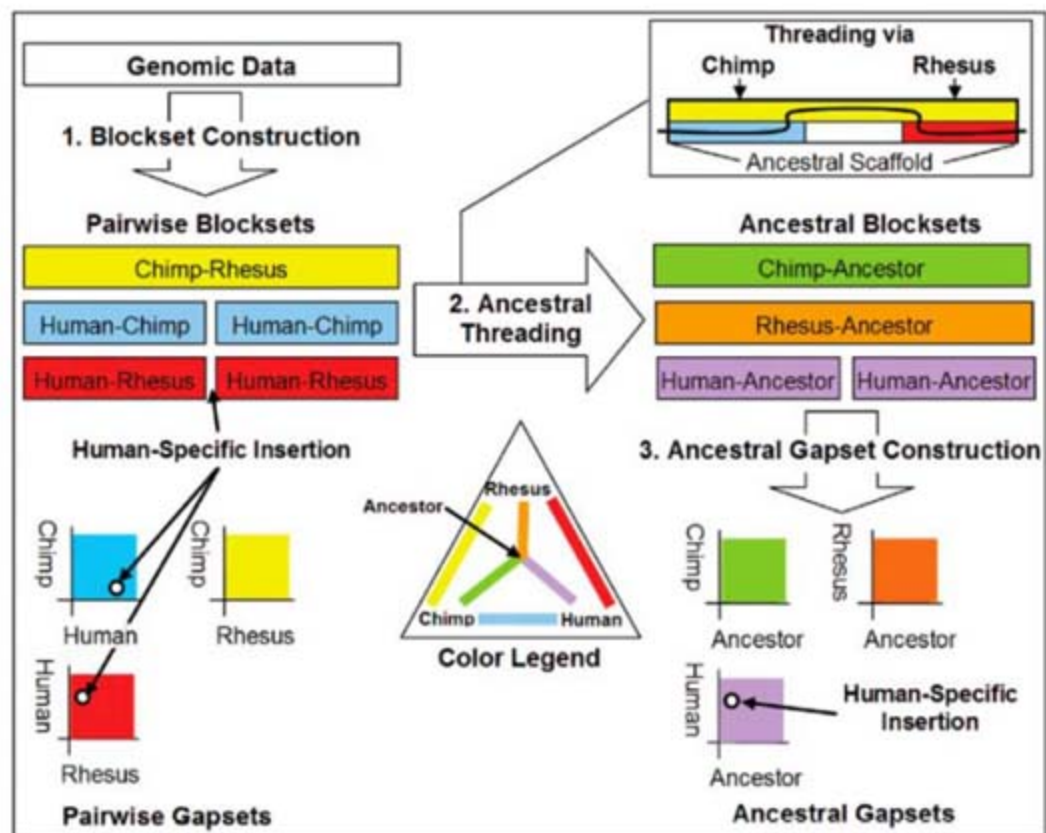

Fig. 1. Detection by genomic triangulation of a human-specific breakpoint induced by an insertion. In the example shown (left), blockset construction revealed one long conserved block for a chimp-rhesus comparison. However, blocks from human-chimp and human-rhesus blocksets align except for a gap, which suggests that there has been a human-specific insertion. This is further illustrated by the pairwise gapsets. Pairwise blocksets and gapsets between extant species are represented by primary colors (red, blue, and yellow) and those between an extant species and the ancestor are represented by complementary colors (purple, green, and orange). Gapsets are represented as gap maps, with white circles representing gaps and circle coordinates indicating sizes of the gaps in specific genomes. Gapsets on the left indicate breakpoints in human-chimp and human-rhesus pairwise comparisons due to the human breakpoint. Ancestral gapsets on the right indicate the breakpoint between human and the ancestor.

Fig. 2. Human-specific breakpoints. (**A**) Chromosome ideograms of 130 human-specific breakpoints detected by genomic triangulation, with breakpoint counts by chromosome. (**B**) Gap map with ancestor gap lengths on the x axis and human gap lengths on the y axis. Only gaps greater than 10 kb in at least one of the genomes and smaller than 4 Mb were considered. Some insertions occur to the right of the $y = x$ axis because visual inspection revealed inaccuracies in gap sizing, providing evidence for human insertion. (**C**) Size distribution of detected human-specific insertions. The percent identity of segmental duplications can indicate the approximate age of duplication events.
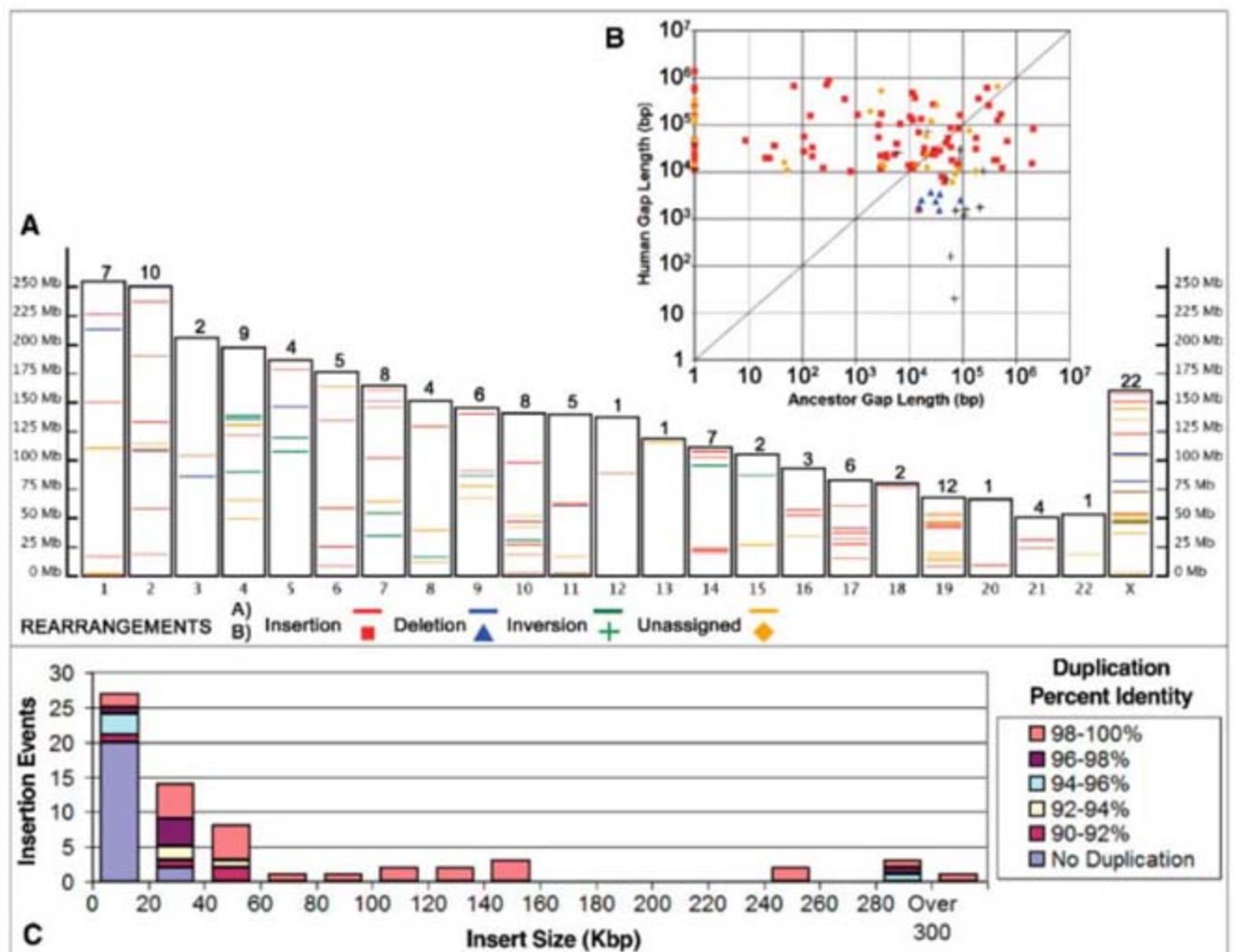
First, because the fragments of an ancestral genome cannot be sequenced directly, they are inferred abductively in the blockset construction step. In other words, each block of similarity is explained by hypothesizing that a corresponding fragment was present in the ancestral genome. Second, the fragments are not assembled into ancestral scaffolds abductively as in genome assembly, where read fragments with similar sequences at their ends are assembled into a contig under the hypothetical assumption that their similarity is due to their origin from overlapping genomic loci. Instead, the ancestral threading step infers the ancestral genome structure deductively from block overlaps in assembled genomes of extant species. In contrast to the contiguous ancestral region (CAR)–building method (*26*), which relies solely on alignment nets between assembled genomes, genomic triangulation integrates assembly anchoring data with clone end sequencing and physical map information, thereby reducing problems associated with draft-quality assemblies. CAR can detect inversions, translocations, fissions, and fusions larger than 50 kb in length but is insensitive to insertions and deletions detectable by genomic triangulation.

The genomic triangulation method introduces the gapset, a key concept complementary to that of a blockset (*12*). Gapsets provide a means of

tracking breakpoints, which cannot be adequately expressed using blocksets. A gapset operationally defines breakpoints under minimal assumptions, thus helping to integrate genome information from multiple sources. Ancestral gapsets localize breakpoints both by position, by flanking blocks within a genome, and phylogenetically, by assigning them to a specific branch of a phylogenetic tree, thus providing maximal information for the reconstruction of evolutionary changes in genome structure.

References and Notes

1. J. J. Yunis, O. Prakash, *Science* **215**, 1525 (1982).
2. J. Wienberg, *Cytogenet. Genome Res.* **108**, 139 (2005).
3. M. C. Popesco et al., *Science* **313**, 1304 (2006).
4. A. Fortna et al., *PLoS Biol.* **2**, E207 (2004).
5. G. M. Wilson et al., *Genome Res.* **16**, 173 (2006).
6. V. Goidts et al., *Hum. Genet.* **120**, 270 (2006).
7. T. L. Newman et al., *Genome Res.* **15**, 1344 (2005).
8. The Chimpanzee Sequencing and Analysis Consortium, *Nature* **437**, 69 (2005).
9. M. Rocchi, N. Archidiacono, R. Stanyon, *Genome Res.* **16**, 1441 (2006).
10. L. Froenicke et al., *Genome Res.* **16**, 306 (2006).
11. G. Bourque, G. Tesler, P. A. Pevzner, *Genome Res.* **16**, 311 (2006).
12. M. Blanchette et al., *Genome Res.* **14**, 708 (2004).
13. The Rhesus Macaque Genome Sequencing and Analysis Consortium, *Science* **316**, 222 (2007).
14. K. J. Kalafus, A. R. Jackson, A. Milosavljevic, *Genome Res.* **14**, 672 (2004).
15. W. J. Kent, R. Baertsch, A. Hinrichs, W. Miller, D. Haussler, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11484 (2003).
16. A. Milosavljevic et al., *Genome Res.* **15**, 292 (2005).
17. J. Rogers et al., *Genomics* **87**, 30 (2006).
18. W. J. Murphy et al., *Genomics* **86**, 383 (2005).
19. J. A. Bailey, A. M. Yavor, H. F. Massa, B. J. Trask, E. E. Eichler, *Genome Res.* **11**, 1005 (2001).
20. Z. Cheng et al., *Nature* **437**, 88 (2005).
21. M. T. Ross et al., *Nature* **434**, 325 (2005).
22. N. Patterson, D. J. Richter, S. Gnerre, E. S. Lander, D. Reich, *Nature* **441**, 1103 (2006).
23. N. H. Barton, *Curr. Biol.* **16**, R647 (2006).
24. The Centre for Applied Genomics, http://projects.tcag.ca/variation/).
25. G. Myers, *Comput. Sci. Eng.* **1**, 33 (1999).
26. J. Ma et al., *Genome Res.* **16**, 1557 (2006).

REPORT

Mobile DNA in Old World Monkeys: A Glimpse Through the Rhesus Macaque Genome

Kyudong Han,[1]* Miriam K. Konkel,[1]* Jinchuan Xing,[1]*† Hui Wang,[1]* Jungnam Lee,[1] Thomas J. Meyer,[1] Charles T. Huang,[1] Erin Sandifer,[1] Kristi Hebert,[1] Erin W. Barnes,[1] Robert Hubley,[2] Webb Miller,[3] Arian F. A. Smit,[2] Brygg Ullmer,[4] Mark A. Batzer[1]‡

The completion of the draft sequence of the rhesus macaque genome allowed us to study the genomic composition and evolution of transposable elements in this representative of the Old World monkey lineage, a group of diverse primates closely related to humans. The L1 family of long interspersed elements appears to have evolved as a single lineage, and *Alu* elements have evolved into four currently active lineages. We also found evidence of elevated horizontal transmissions of retroviruses and the absence of DNA transposon activity in the Old World monkey lineage. In addition, ~100 precursors of composite SVA (short interspersed element, variable number of tandem repeat, and *Alu*) elements were identified, with the majority being shared by the common ancestor of humans and rhesus macaques. Mobile elements compose roughly 50% of primate genomes, and our findings illustrate their diversity and strong influence on genome evolution between closely related species.

Old World monkeys (OWMs) represent one of the most closely related primate groups to humans. The rhesus macaques (*Macaca mulatta*), along with other OWMs, have been extensively used in biomedical studies (*1*). An improved understanding of their genomic architecture could hold important implications for medicine, evolutionary understanding, and beyond. Similar to the human and chimpanzee genomes, roughly 50% of the rhesus macaque genome consists of various repetitive sequences (*2–4*). The majority of these repeats are mobile elements, which can be divided into class I DNA transposons (*5*) and class II retrotransposons (*6*). Related transposable elements are further categorized into families, with each family further classified into subfamilies on the basis of their sequence relationships. The insertion of mobile elements can alter gene expression (*7*), generate genomic deletions (*8*), and even create new genes and gene families (*9*). Existing repetitive elements can also mediate recombinations between similar elements at different genomic locations (ectopic recombination) (*10*). In addition, the GC-rich nature of certain mobile elements {e.g., *Alu* and SVA [short interspersed element (SINE), variable number of tandem repeat (VNTR), and *Alu*] elements} can introduce new GC islands through their insertion (*3*). Despite the overall similarity in retrotransposon mobilization activity in the OWM and hominoid (human and ape) lineages, mobile elements have continued to evolve independently in both lineages. Close examination of the overall mobile-element composition in OWMs, with the rhesus macaque genome used as a reference, allows an understanding of their lineage-specific expansion and illustrates their overall contribution to genome evolution.

Without any detected lineage-specific copies, DNA transposons, which mobilize through a cut-and-paste mechanism, appear to have been inactive in the rhesus macaque lineage since their speciation from humans. The paucity of DNA transposon mobilization in mammals, and in amniotes in general, is noteworthy by comparison with other organisms (e.g., plants) and may result from the relative difficulty in horizontal transfer into animals' germ lines (*11*).

Similar to the human genome, the rhesus macaque genome contains over half a million recognizable copies of endogenous retroviruses (ERVs) and their nonautonomous derivatives, with the great majority being present or fixed before the hominoid-OWM split (*12*). We found evidence for at least eight instances of horizontal transmission of ERVs in the OWM lineage resulting in 2750 extant copies (table S1 and SOM Text). This is much higher than in the human lineage, where there is evidence for only one or two invading elements leaving fewer than

10 extant copies (*13*). Five of the eight horizontally transmitted ERVs belong to class I retroviruses, and the remaining three belong to class II retroviruses (shown in red letters in Fig. 1). Apart from these new invasions, at least seven ERV families already entered the genome before the hominoid-OWM split and remained active afterward. There are over 3500 copies of these ERV subfamilies in the OWM lineage, similar to the number of lineage-specific ERV copies in humans.

The L1PA (primate A) family of long interspersed elements (LINES) represents the dominant active L1 lineage throughout primate evolution. In our analysis, L1PA5 was the most commonly recovered L1 subfamily, and ~19,000 L1PA5 elements specific to the OWM lineage were identified in the rhesus macaque genome. Most of these elements represent insertion events that occurred along the OWM lineage leading to rhesus macaques and are therefore present in multiple OWM species (fig. S2). A total of 32 OWM-specific L1 subfamilies were identified with the use of diagnostic substitutions present in these elements (table S2). To investigate the relationship of L1s, we constructed a median-joining network with their consensus sequences (Fig. 2 and SOM Text) and estimated the age of each subfamily (table S2). The network results indicated that the OWM-specific L1 lineage rooted with the L1PA6 consensus sequence, and several lineages roughly followed a sequential order, with little overlap in their amplification period. The sequential evolution of L1 elements appears to follow a general trend seen in mammalian L1s (*14*) and may result from amplification competition between two distinct L1 lineages (*15*). Altogether, we identified nine putative retrotransposition-competent L1s in the rhesus macaque genome, and they belonged to the L1CER-3 or L1CER-4 subfamilies; each L1 subfamily name is identified by "CER" (which stands for Cercopithecidae, indicating the origin of the consensus sequence) and an Arabic numeral indicating its lineage (*12*). Nine was a considerably lower number of potentially active L1 elements than that in the human genome, which has 80 to 100 active copies (*16*). Nevertheless, it is likely that additional retrotransposition-competent L1 elements will be recovered in more refined drafts of the rhesus macaque genome.

Retrotransposon-mediated DNA sequence transduction is a process whereby a retrotransposon carries a flanking genomic sequence during its mobilization that can result in exon or gene duplication (*17*). Three L1 elements with 5′ transduced exon–derived sequences were identified in the rhesus macaque genome. Moreover, detailed analysis indicated that one of the three insertions occurred in an exon of another gene (table S3 and SOM Text). These three events empirically demonstrate that exon-derived sequences can be transferred via 5′ L1–mediated transduction within primate genomes and that 5′ transduction constitutes a second mechanism of retrotransposon-mediated "exon shuffling."
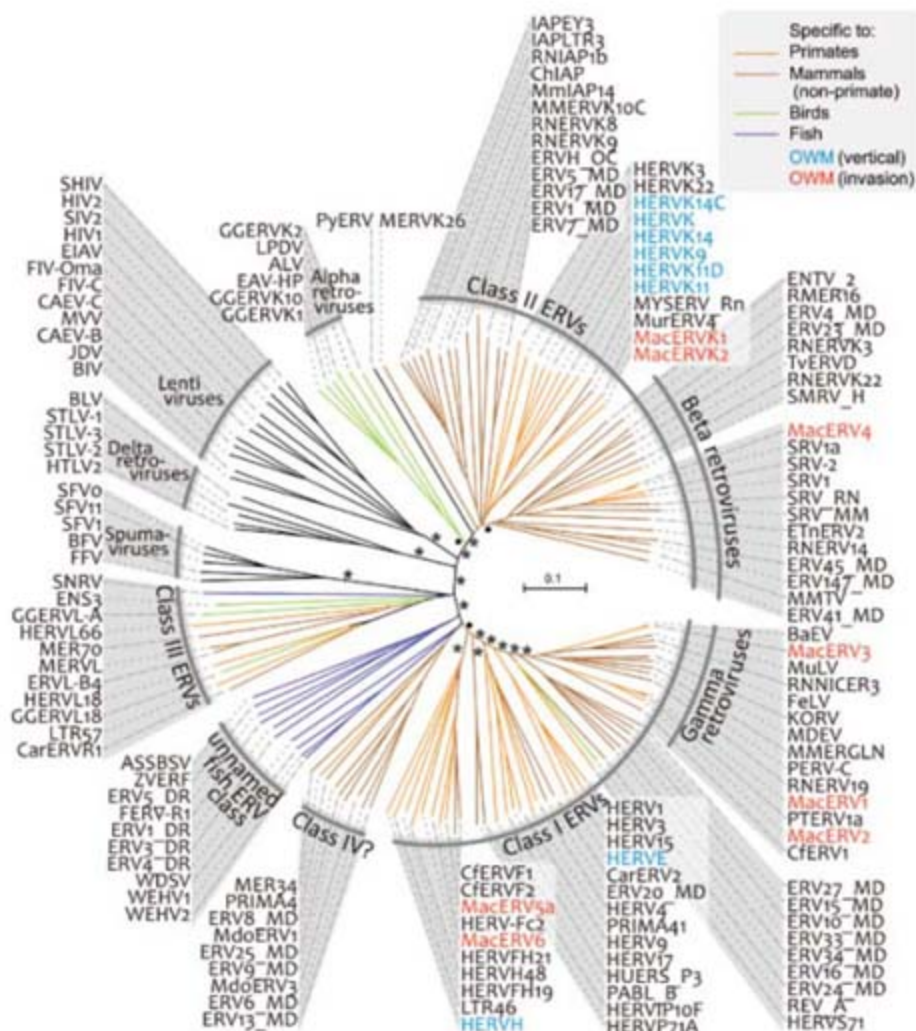
Fig. 1. Phylogenetic tree of retroviruses based on full-length Pol proteins. Common infectious retroviruses and endogenous retroviruses, present in fish, birds, mammals (nonprimate), and primates, were included in the analysis. Color identifications for each group are shown in the upper right corner. Asterisks and circles show deep-rooted branches with >95 and >75% bootstrap values, respectively. The ERVs identified in this study that invaded the OWM genome horizontally (i.e., through external germline infection) are indicated with red letters. For all ERVs shown in blue letters, the original insertion occurred in the common ancestor of humans and rhesus macaques (i.e., vertically) and is present in both genomes. All ERVs indicated with blue letters also generated new insertions in the OWM lineage. The scale bar indicates 10% divergence in the amino acid sequence.

Fig. 2. Median-joining network of OWM-specific L1 subfamilies. Subfamilies are represented by circles, with the circle size symbolizing the relative size of each subfamily. The length of the lines corresponds to the number of substitutions. The scale of a single substitution is shown in the upper left corner. Broken lines indicate segments not drawn to scale. Gray circles represent the subfamilies belonging to the **L1CER-3** lineage, which include an 18—base pair (bp) duplication in their 3′ untranslated region (3′UTR), and green-edged circles contain intact full-length L1 elements. The dashed line and red arrow represent two alternative pathways for the origin of the **L1CER-4** subfamily. The subfamilies in the blue and pink ovals share the same diagnostic mutations but do not share the 18-bp duplication. My, million years.

Alu elements are the most successful SINEs in primate genomes (*18*), and ~110,000 *Alu* insertions are specific to OWMs. Fourteen different OWM lineage-specific *Alu*Y subfamilies fell into four lineages, shown in a median-joining network analysis (Fig. 3), and were identified with estimated copy numbers (table S4). All subfamilies were estimated to have originated after the hominoid-OWM divergence and were congruent with our phylogenetic analyses showing that all of these *Alu* subfamilies were restricted to OWMs (SOM Text). The simultaneous retrotransposition activity of multiple *Alu* subfamilies is similar to that in the human genome, and the activity of multiple "source genes" may have contributed to the amplification success of *Alu* elements despite their reliance on L1 enzymatic machinery for mobilization (*19*).

About 100 precursors of SVA were identified in the rhesus macaque genome. The variable number of tandem repeat (VNTR) regions of these elements share >90% identity with the VNTR unit in hominoid SVA elements (*20*), although they have no sequence homology with other components of SVA elements. Thus, these elements appear to have contributed a portion of the genetic material required to form the SVA composite retrotransposon family in hominoids. The majority of these elements are shared between human and rhesus macaque, indicating that these elements were active before the divergence of hominoids and OWMs. The low number of lineage-specific elements (~20 in the OWM lineage) suggests a very low retrotransposition rate of SVA precursor elements over the past 25 million years.

Composing nearly half of all sequenced primate genomes, mobile elements—especially retrotransposons—are major components of genomic variation and a driving force of primate evolution. Although the overall number of mobile elements is similar in the human, chimpanzee, and rhesus macaque genomes (*2–4*), a large fraction of the elements inserted independently into different locations within each genome and thus shaped the genomes differently (*21*). Whereas most retrotransposon insertions remain neutral in the genome, many insertions can have deleterious effects of varying severity. Mobile elements can cause genetic diseases not only by direct gene disruption or by the deletion of exonic sequence upon insertion but also by mediating subsequent recombination between existing retrotransposons. Indeed, more than 118 human genetic disorders are caused by retrotransposons, including hemophilia B, breast cancers, and congenital muscular dystrophy [see (*22*) and (*23*) for reviews]; they are likely to have a similar impact on the rhesus genome. Yet, retrotransposons are also responsible for creating a variety of genomic novelties. They are involved in mediating gene duplication, exon shuffling, and RNA-editing–mediated exonization (*9, 17, 24*). All these mechanisms can contribute to new gene formation, as well as potentially altering DNA methylation patterns and contributing to X chro-
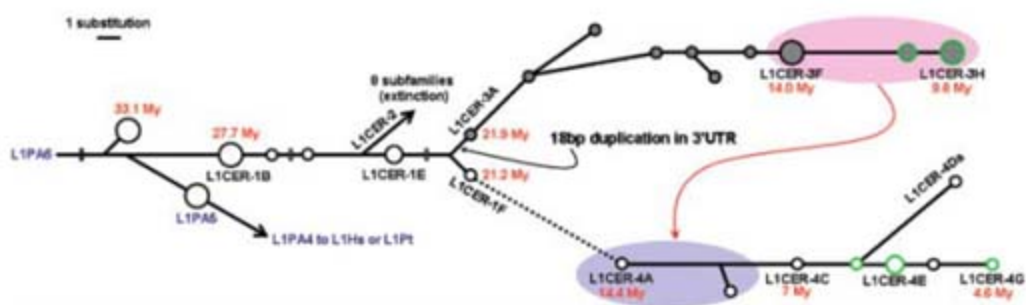
Fig. 3. Median-joining network of OWM-specific *Alu* subfamilies. Subfamilies are represented by circles. The length of the lines corresponds to the number of substitutions, and the scale of a single substitution is shown in the upper left corner. Broken lines indicate segments not drawn to scale. Gray circles represent all subfamilies belonging to the *Alu*YRb lineage containing a 12-bp deletion. Red-edged circles denote the youngest *Alu* subfamily within each lineage, and the blue-edged circle indicates the *Alu*Y subfamily consensus sequence.
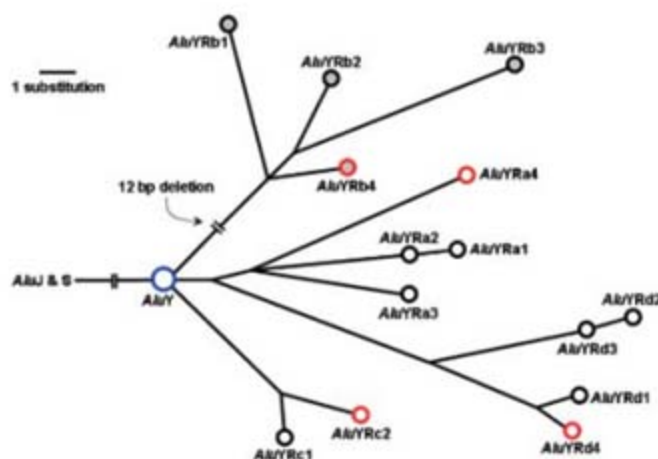
mosome inactivation in females (*25, 26*). In addition, retrotransposons provide highly valuable genetic systems for primate population and phylogenetic studies, because they have a known ancestral (i.e., insertion-absent) state, and the chance that the same type of element would integrate at precisely the same location in multiple individuals is essentially zero (i.e., the insertions are identical by descent) (*27, 28*). Altogether, understanding the mobile-element landscape in primates is not only important for biologists but also crucial for biomedical researchers using primate animal models.

References and Notes

1. H. E. Carlsson, S. J. Schapiro, I. Farah, J. Hau, *Am. J. Primatol.* **63**, 225 (2004).
2. Chimpanzee Sequencing and Analysis Consortium, *Nature* **437**, 69 (2005).
3. E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
4. Rhesus Macaque Genome Sequencing and Analysis Consortium, *Science* **316**, 222 (2007).
5. A. F. Smit, *Curr. Opin. Genet. Dev.* **6**, 743 (1996).
6. P. L. Deininger, M. A. Batzer, *Genome Res.* **12**, 1455 (2002).
7. P. L. Deininger, J. V. Moran, M. A. Batzer, H. H. Kazazian Jr., *Curr. Opin. Genet. Dev.* **13**, 651 (2003).
8. K. Han *et al.*, *Nucleic Acids Res.* **33**, 4040 (2005).
9. J. Xing *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17608 (2006).
10. S. K. Sen *et al.*, *Am. J. Hum. Genet.* **79**, 41 (2006).
11. A. F. Smit, *Curr. Opin. Genet. Dev.* **9**, 657 (1999).
12. Materials and methods are available as supporting material on *Science* Online.
13. L. Benit, A. Calteau, T. Heidmann, *Virology* **312**, 159 (2003).
14. A. V. Furano, D. D. Duvernell, S. Boissinot, *Trends Genet.* **20**, 9 (2004).
15. H. Khan, A. Smit, S. Boissinot, *Genome Res.* **16**, 78 (2006).
16. B. Brouha *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5280 (2003).
17. J. V. Moran, R. J. DeBerardinis, H. H. Kazazian Jr., *Science* **283**, 1530 (1999).
18. M. A. Batzer, P. L. Deininger, *Nat. Rev. Genet.* **3**, 370 (2002).
19. R. Cordaux, D. J. Hedges, M. A. Batzer, *Trends Genet.* **20**, 464 (2004).
20. H. Wang *et al.*, *J. Mol. Biol.* **354**, 994 (2005).
21. R. E. Mills *et al.*, *Am. J. Hum. Genet.* **78**, 671 (2006).
22. P. A. Callinan, M. A. Batzer, in *Genome Dynamics*, J. N. Volff, Ed. (Karger, Basel, Switzerland, 2006), vol. 1, pp. 104–115.
23. J. M. Chen, C. Ferec, D. N. Cooper, *J. Biomed. Biotechnol.* **2006**, 56182 (2006).
24. G. Lev-Maor *et al.*, *Genome Biol.* **8**, R29 (2007).
25. J. A. Bailey, L. Carrel, A. Chakravarti, E. E. Eichler, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6634 (2000).
26. R. S. Hansen, *Hum. Mol. Genet.* **12**, 2559 (2003).
27. A. M. Shedlock, K. Takahashi, N. Okada, *Trends Ecol. Evol.* **19**, 545 (2004).
28. D. A. Ray, J. Xing, A. H. Salem, M. A. Batzer, *Syst. Biol.* **55**, 928 (2006).
29. Thanks to the RMGSAC for the rhesus macaque genome sequence and to S. Brandt, W. Scullin, and S. White for computational support. This project was facilitated in part by high-performance computing allocations from Louisiana State University CCT and supported by the NSF grants BCS-0218338 (M.A.B.) and EPS-0346411 (M.A.B.), NIH GM59290 (M.A.B.), and the State of Louisiana Board of Regents Support Fund (M.A.B.).

REPORT

Demographic Histories and Patterns of Linkage Disequilibrium in Chinese and Indian Rhesus Macaques

Ryan D. Hernandez,[1] Melissa J. Hubisz,[2] David A. Wheeler,[3] David G. Smith,[4,5] Betsy Ferguson,[6,7] Jeffrey Rogers,[8] Lynne Nazareth,[3] Amit Indap,[1] Traci Bourquin,[3] John McPherson,[3] Donna Muzny,[3] Richard Gibbs,[3] Rasmus Nielsen,[9] Carlos D. Bustamante[1]*

To understand the demographic history of rhesus macaques (*Macaca mulatta*) and document the extent of linkage disequilibrium (LD) in the genome, we partially resequenced five Encyclopedia of DNA Elements regions in 9 Chinese and 38 captive-born Indian rhesus macaques. Population genetic analyses of the 1467 single-nucleotide polymorphisms discovered suggest that the two populations separated about 162,000 years ago, with the Chinese population tripling in size since then and the Indian population eventually shrinking by a factor of four. Using coalescent simulations, we confirmed that these inferred demographic events explain a much faster decay of LD in Chinese ($r^2 \approx 0.15$ at 10 kilobases) versus Indian ($r^2 \approx 0.52$ at 10 kilobases) macaque populations.

Rhesus macaques (*Macaca mulatta*) and humans shared a most recent common ancestor (MRCA) ~25 million years ago (Ma), and our genomes differ at <7% of nucleotide bases (*1*). Rhesus and humans, therefore, share a large number of fundamental biological characteristics, including many underlying genetic and physiological processes that lead to disease. For this reason, rhesus macaques have become a model organism for vaccine research (*2, 3*), as well as

studies of normal human physiology and disease. Although previous studies of genetic variation in rhesus have described >300 microsatellite polymorphisms (*4, 5*), identifying specific genetic risk factors for disease requires a much greater resolution of genetic variation across the genome.

The current geographic range of rhesus macaques is larger than any other nonhuman primate, stretching from western India and Pakistan to the eastern shores of China (Fig. 1). Fossil records suggest that the genus *Macaca* originated

[1]Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14850, USA. [2]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA. [3]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA. [4]Department of Anthropology, Davis, CA, USA. [5]California National Primate Research Center, Davis, CA, USA. [6]Genetics Research and Informatics Program, Oregon National Primate Research Center, Oregon Health and Sciences University, Beaverton, OR 97006, USA. [7]Washington National Primate Research Center, University of Washington, Seattle, WA 98195, USA. [8]Department of Genetics, Southwest Foundation for Biomedical Research, and Southwest National Primate Research Center, San Antonio, TX 78227, USA. [9]Center for Comparative Genomics, Department of Biology, University of Copenhagen, Universitetsparken 15, 2100 Kbh Ø, Denmark.

*To whom correspondence should be addressed. E-mail: cdb28@cornell.edu

in northern Africa approximately 5.5 Ma, followed by migration through the Middle East and into northern India by ~3 Ma (6). By ~2 Ma, macaques had traversed most of China and reached the Indonesian archipelago, where the putative ancestral species of rhesus macaque, *M. fascicularis*, is thought to have originated (6, 7).

Previous studies of mitochondrial DNA (8), major histocompatibility complex (MHC) alleles (9), and single-nucleotide polymorphisms (SNPs)

in gene-linked regions (10) suggest moderate levels of genetic differentiation between captive-born Indian and Chinese rhesus populations. Developing a more thorough understanding of genetic variation within and between these two populations has important implications for biomedical research. For example, when infected with the simian immunodeficiency virus, animals from Chinese populations develop AIDS-like symptoms more slowly than animals from Indian populations (3).

We have identified 1476 SNPs by sequencing >150 kb of DNA across five Encyclopedia of DNA Elements (ENCODE) (11–13) regions located on separate autosomal chromosomes in nine captive-born from wild-caught Chinese and 38 captive-born Indian rhesus macaques. The Chinese animals derive from three distinct geographical sites, whereas the Indian animals came from three different colonies in the United States (Fig. 1). Individuals were chosen to represent rhesus macaque populations that are currently being studied by the international community and to minimize relatedness in the sample [with most individuals in the study being unrelated back to the founding of the colony into which they were born, and none having a shared grandparent (13)]. In our sample of 1476 SNPs discovered, only 486 (33%) were shared across both populations, whereas 604 were found only in the Chinese population (61% of 1090 SNPs observed) and 386 were found only in the Indian population (39% of 872 SNPs observed). The frequency distribution of derived mutations across SNPs [using DNA sequence from the ENCODE project for baboon, *Papio cynocephalus anubis*, to infer the putative ancestral state (13)] shows that the Chinese population harbors an excess of rare SNPs relative to a population of constant size, whereas the Indian population has too few rare and too many intermediate- and high-frequency–derived SNPs (Fig. 2A). The observed disparity in SNP density (7.25 SNPs per kb for Chinese versus 5.8 SNPs per kb for Indian) in the two populations suggests that the effective size of the Chinese population is much larger than the Indian population, given that the Indian sample size is four times as large as that of the Chinese.

We observed a moderate level of population structure between the Indian and Chinese samples, as measured by Wright's F_{ST} statistic (average $F_{ST} = 0.14$; SD = 0.11; range = –0.024 to 0.645) (Fig. 3A). Furthermore, the Bayesian clustering program STRUCTURE (14) clearly separates Chinese and Indian individuals when assuming two clusters (Fig. 3B), and considering more clusters does not significantly improve the fit of the model. We found only one Chinese individual with a marginal amount of Indian ancestry (8.5%, sampled from Suzhou) and eight Indian individuals with more than 5% Chinese ancestry [max 16.8%, including animals from all three primate centers (13)]. These low levels of admixture suggest that recurrent migration between the populations has been minimal. Moreover, the two populations were clearly distinguished by principal components analysis (15) along the first two axes of variation (Fig. 3C). Interestingly, the second component also separates one Chinese individual (sampled from Suzhou) from the others, which suggests that further population substructure may exist. Although this individual is not differentiated from other Chinese-origin animals in the STRUCTURE analysis, it may, nonetheless, harbor alleles from an unsampled

Fig. 1. The current geographic range of rhesus macaques [green, redrawn from (20)] with the inferred demographic history and the sample locations superimposed. The geographic location of the MRCA is based on (4).

Fig. 2. (A) The marginal frequency spectrum of derived mutations for each population (shown as expected proportions in a subsample of 10 chromosomes by integrating over possible configurations of observed and missing data, with the total number of SNPs in parentheses) and the expected distribution under the standard neutral model (SNM) of constant size. (B) A "topographical map" of the joint site-frequency spectrum for the two populations, with darker tones representing frequency pairs with few SNPs, and lighter tones representing frequency pairs with many SNPs.

The Rhesus Macaque Genome

Fig. 3. (**A**) The distribution of F$_{ST}$ between Indian and Chinese rhesus, calculated with the average pairwise-difference across each nonoverlapping window (*13*). (**B**) STRUCTURE results. Individuals are represented by vertical lines, and sorted by their amount of Chinese ancestry (black vertical line separates animals with Indian and Chinese origins). Colors correspond to the proportion of an individual's ancestry attributable to a given population (blue, Indian; red, Chinese). (**C**) Principal component 1 (PC1) and PC2 separate Indian from Chinese individuals. PC2 also isolates a single Chinese individual [corresponding to an individual sampled from Suzhou and shown as the fourth individual from the right in (**B**)].
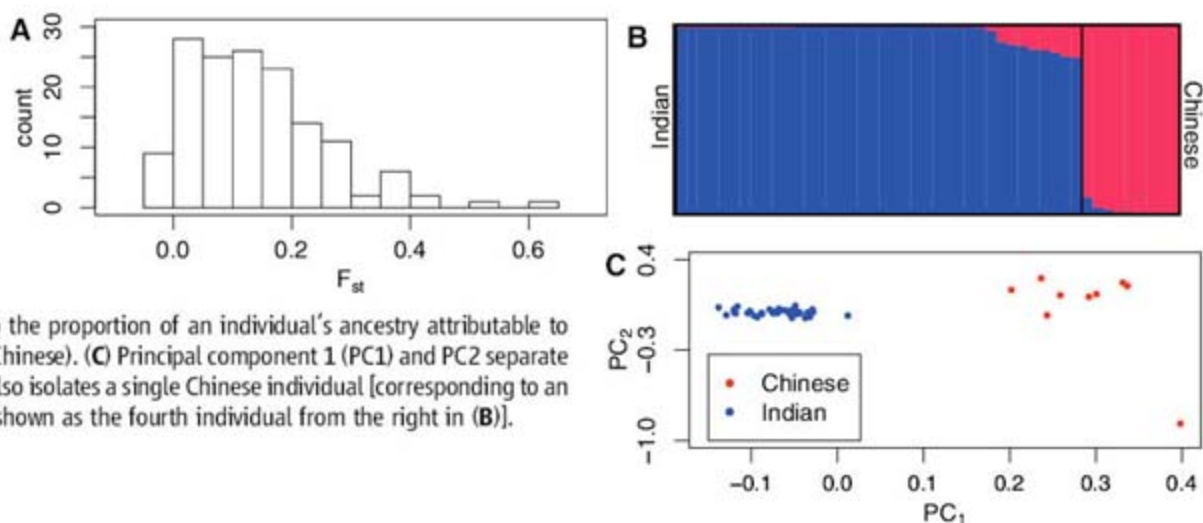
Fig. 4. The decay of LD for Indian and Chinese rhesus macaques versus European and African humans (*n* = 9 for all samples), along with the decay of LD for 1000 neutral simulations of our inferred demographic history for rhesus macaque. Human data are from three ENCODE regions orthologous to the rhesus data (*13, 21*).
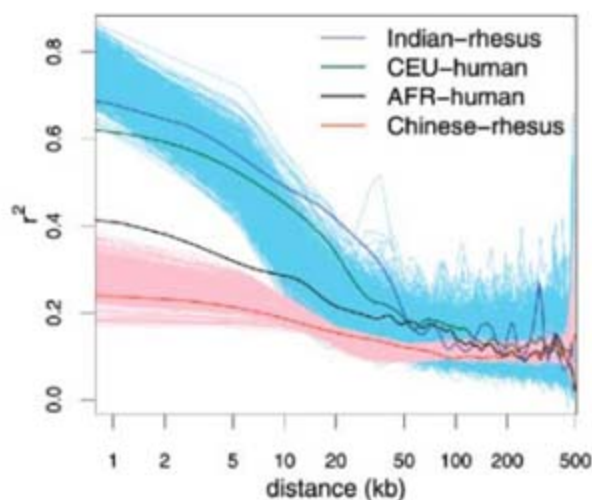
Chinese subpopulation (i.e., the two wild-caught parents may be from different subpopulations).

Using maximum likelihood under the assumption that the animals in this study form a random sample from their respective population (*13*), we fit a two-population demographic model to the joint distribution of SNP frequencies, or site-frequency spectrum, shown in Fig. 2B. Our model suggests that the Chinese population expanded by a factor of 3.3 and separated from the Indian population ~162 thousand years ago (ka) (95% confidence interval, CI = 183 to 132 ka). After separating, the Indian population maintained its ancestral population size until ~51 ka CI = 72 to 21 ka)], when it was reduced by a factor of 4.3. The population genetic model, although a very simplistic approximation to the rich and complex history of the species, fits the data well, as indicated by a goodness-of-fit test (*P* = 0.133). Coalescent simulations (*13*) on the basis of the inferred demographic history for Indian and Chinese rhesus macaques suggest that the MRCA of the two populations lived ~1.94 Ma (SE 14 Ky). This estimate places the MRCA of rhesus near the divergence time from *M. fascicularis*, inferred from mitochondrial DNA to be 1.83 to 5 Ma (*16, 17*). Moreover, our simulations suggest that the effective size of the ancestral population of rhesus

macaques was ~73,070 (SE 231) individuals, implying that the current effective size of the Chinese population is ~239,704, whereas the Indian population is estimated to be ~17,014.

The recent demographic events that caused these differences in effective population sizes of Indian and Chinese rhesus macaques have also had a large impact on linkage disequilibrium (LD). To quantify the extent of LD in Indian and Chinese rhesus macaques, we measured the correlation coefficient (r^2) of alleles from frequency-matched SNPs (*13, 18*). Figure 4 shows substantial differences between the Indian and Chinese rhesus macaque populations, which are more extreme than the patterns observed among humans. For example, within the Indian rhesus population, LD extends much further than LD observed for European humans, whereas the Chinese rhesus population shows little LD, even for SNPs that are physically very close. Coalescent simulations (*13*) show that the observed patterns of LD are consistent with our inferred demographic history of this species (shown in Fig. 4 as light blue and pink curves for Indian and Chinese rhesus, respectively). However, LD in the Indian population extends slightly further than expected. This observation may be consistent with recent admixture with a Burmese rhesus population not sam-

pled in this study (*8*), because admixture between populations with allele frequency differences is known to generate long-range LD.

In this study, we analyzed noncoding data in rhesus macaques to characterize their underlying demographic history and to quantify the extent of LD relative to humans. The genetic differences that we have observed between Indian and Chinese rhesus macaques are consistent with a recent report on the distribution of SNPs in these populations (*10*), as well as previous studies of protein coding, microsatellite STR (short tandem repeat), MHC loci, and mitochondrial and Y-chromosome DNA haplotypes (*8*). Without samples from wild-caught Indian rhesus monkeys, however, these data must be regarded as estimates, because they may reflect a sampling bias toward those macaques that are available for study in the United States as a result of international restrictions on exportation of primates.

Extending these studies to whole-genome association mapping in captive-born animals could be fruitful for identifying genes involved in human diseases. On the basis of the patterns of LD that we have observed, such an association study would likely require many fewer markers to identify common disease-causing variants in rhesus macaques than in humans. Because LD in captive Indian rhesus macaque populations extends much further than in humans, a SNP map with roughly 1 SNP every 35 kb (82,000 SNPs total) would suffice to achieve the same threshold (r^2 = 0.4) as a marker every 6 kb in humans (*13, 19*). Furthermore, because LD decays much faster in Chinese rhesus monkeys than in humans, Chinese macaques provide an ideal platform for localizing mutations that are difficult to map in either Indian macaques or humans as a result of extensive LD among candidate mutations in a particular region.

References and Notes
1. Rhesus Macaque Genome Sequencing and Analysis Consortium, *Science* **316**, 222 (2007).
2. R. A. Weiss, *Nature* **410**, 1035 (2001).
3. B. Ling *et al.*, *AIDS* **16**, 1489 (2002).
4. J. Rogers *et al.*, *Genomics* **87**, 30 (2006).

5. M. Raveendran et al., Genomics 88, 706 (2006).
6. E. Delson, in The Macaques: Studies in Ecology, Behavior, and Evolution, D. D. Lindburg, Ed. (van Nostrand Rheinhold, New York, 1980), pp. 10–30.
7. C. Abegg, B. Thierry, Biol. J. Linn. Soc. 75, 555 (2002).
8. D. G. Smith, J. McDonough, Am. J. Primatol. 65, 1 (2005).
9. J. Viray, B. Rolfs, D. G. Smith, Comp. Med. 51, 555 (2001).
10. B. Ferguson et al., BMC Genom. 8, 43 (2007).
11. ENCODE Project Consortium, Science 306, 636 (2004).
12. ENCODE regions were chosen because they have been widely studied across several mammals, including rhesus and baboon.
13. Materials and methods are available as supporting material on Science Online.

14. D. Falush, M. Stephens, J. K. Pritchard, Genetics 164, 1567 (2003).
15. A. L. Price et al., Nat. Genet. 38, 904 (2006).
16. K. Hayasaka, K. Fujii, S. Horai, Mol. Biol. Evol. 13, 1044 (1996).
17. J. C. Morales, D. J. Melnick, J. Hum. Evol. 34, 1 (1998).
18. M. A. Eberle, M. J. Rieder, L. Kruglyak, D. A. Nickerson, PLoS Genet. 2, 1319 (2006).
19. L. Kruglyak, Nat. Genet. 22, 139 (1999).
20. J. Fooden, in The Macaques: Studies in Ecology, Behavior, and Evolution, D. D. Lindburg, Ed. (van Nostrand Rheinhold, New York, 1980), pp. 1–9.
21. HapMap, Nature 437, 1299 (2005).

REPORT

Evolutionary Formation of New Centromeres in Macaque

Mario Ventura,[1]* Francesca Antonacci,[1]* Maria Francesca Cardone,[1] Roscoe Stanyon,[2] Pietro D'Addabbo,[1] Angelo Cellamare,[1] L. James Sprague,[3] Evan E. Eichler,[3] Nicoletta Archidiacono,[1] Mariano Rocchi[1]†

A systematic fluorescence in situ hybridization comparison of macaque and human synteny organization disclosed five additional macaque evolutionary new centromeres (ENCs) for a total of nine ENCs. To understand the dynamics of ENC formation and progression, we compared the ENC of macaque chromosome 4 with the human orthologous region, at 6q24.3, that conserves the ancestral genomic organization. A 250-kilobase segment was extensively duplicated around the macaque centromere. These duplications were strictly intrachromosomal. Our results suggest that novel centromeres may trigger only local duplication activity and that the absence of genes in the seeding region may have been important in ENC maintenance and progression.

Evolutionary new centromeres (ENCs) can appear during evolution in a novel chromosomal region with concomitant inactivation of the old centromere. The new centromere then becomes fixed in the species while inevitably progressing toward the complexity typical of a mature mammalian centromere, with intra- and interchromosomal pericentromeric segmental duplications and a large core of satellite DNA (1). Unambiguous examples of ENCs were initially reported in primates (2) and then described in various other mammalian orders (3). A similar phenomenon, well known from clinical cases, is the mitotic rescue of an acentric chromosomal fragment by the opportunistic de novo emergence of a neocentromere (4). Recently, two cases of neocentromeres in normal individuals with otherwise normal karyotypes were fortuitously discovered (5, 6). These two "in progress" centromeres can be regarded as ENCs at the initial stage, thus reinforcing the opinion that ENCs and clinical neocentromeres are two faces of the same coin. The goal of the research presented here

was to gain insight into the processes and mechanisms of ENC evolution. First, we systematically compared macaque and human synteny organization in search of ENCs. Then, we characterized in detail a macaque ENC and compared it to the orthologous domain in humans, which represents the ancestral genomic structure before ENC seeding.

Multicolor hybridization on rhesus macaque chromosomes [Macaca mulatta (MMU) 2n = 42, where n is the haploid number of chromosomes] of about 500 evenly spaced human bacterial artificial chromosome (BAC) clones

revealed that seven macaque/human homologs (chromosomes 6/5, 8/8, 11/12, 17/13, 19/19, 20/16, and X/X, respectively) were colinear when the position of the centromere was excluded . However, human chromosomes 7/21, 14/15, and 20/22 form syntenic associations as part of three compound macaque chromosomes (3, 7, and 10, respectively). Differences in marker order between macaque and humans were accounted for by 20 chromosome rearrangements. Reiterative fluorescence in situ hybridization (FISH) experiments with additional BAC clones more precisely defined rearrangement breakpoints (table S1). A summary of all results is graphically displayed at www.biologia.uniba.it/macaque. Tables S2 and S3 provide a comprehensive list of the ~600 clones that were used in FISH experiments, from the perspective of both macaque and human chromosomes, respectively.

This comprehensive marker-order comparison revealed that the centromeres of many orthologous chromosomes were embedded in different genomic contexts. To distinguish whether ENC events had occurred in the human or macaque ortholog, or in both, we took into account previous reports that attempted to establish the ancestral form for each chromosome (2, 3, 6–15). (Most of these papers use a different macaque chromosome nomenclature; here we follow the nomenclature used by the macaque genome sequencing consortium. For a comparison, see www.biologia.uniba.it/macaque.) The

[1]Department of Genetics and Microbiology, University of Bari, 70126 Bari, Italy. [2]Department of Animal Biology and Genetics, University of Florence, Florence 50125, Italy. [3]Howard Hughes Medical Institute, Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: rocchi@biologia.uniba.it

Table 1. Macaque chromosomes with neocentromeres. The two noncontiguous positions defining, in human, the ENC of chromosome 1 are due to the colocalization of the ENC with a macaque-specific inversion breakpoint.

MMU (HSA)	Clones	Position of the neocentromere on the human sequence	Reference
1 (1)	RP4-621O15	chr1:226,810,735–226,866,653	Present study
	RP11-572K18	chr1:160,918,751–161,035,790	
2 (3)	RP11-355I21/RP11-418B12	chr3:163,822,353–164,707,155	(6)
4 (6)	RP11-474A9	chr6:145,651,644–145,845,896	(9)
12 (2q)	RP11-343I5/RP11-846E22	chr2:138,659,884–138,908,673	Present study
13 (2p)	RP11-722G17	chr2:86,622,638–86,827,260	Present study
14 (11)	RP11-625D10/RP11-661M13	chr11:5,667,339–6,043,020	(10)
15 (9)	RP11-542K23/RP11-64P14	chr9:124,189,785–124,493,134	Present study
17 (13)	RP11-543A19/RP11-527N12	chr13:61,111,769–62,699,203	(3)
18 (18)	RP11-61D1/RP11-289E15	chr18:50,155,761–50,526,34	Present study

The Rhesus Macaque Genome

results of this analysis confirmed the previously published results and exposed five macaque ENCs (Fig. 1 and Table 1). In total between macaque and human, there are 14 ENCs; 9 ENCs occurred in the macaque lineage [MMU1 (1), 2 (3), 4 (6), 12 (2q), 13 (2p), 14 (11), 15 (9), 17 (13), and 18 (18) (corresponding human chromosomes in parentheses)], and 5 occurred in the human lineage [HSA3 (2), 6 (4), 11 (14), 14 (7a), and 15 (7b) (corresponding macaque chromosomes in parentheses), where HSA denotes *Homo sapiens*]. The newly discovered macaque ENCs were found on MMU1 (1), 12 (2q), 13 (2p), 15 (9), and 18 (18) (corresponding human chromosomes in parentheses). In this context, all macaque centromeres, including the nine ENCs, harbor very large arrays of alpha satellite DNA (*16*) (fig. S1). One possibility is that after their emergence, new macaque centromeres were rapidly stabilized by acquiring alpha satellite DNA.

Human chromosome 6 and the macaque homolog, MMU4, both have ENCs. The ancestral centromere for both species was located at HSA6p22.1 (*9*) (Fig. 2), and the new macaque centromere is located at HSA6q24.3. A comparison of the HSA6q24.3 region [chr6: base pair 139,100,001 to 149,100,000; University of California Santa Cruz (UCSC) March 2006 release] with the orthologous regions of dog, rat, mouse, and opossum genomes, by careful inspection of the specific alignment "Net" in the UCSC genome browser (http://genome.ucsc.edu), showed that a reasonable assumption was that the human region closely resembled the ancestral condition. We reasoned that a detailed comparison of the organization of the MMU4 centromeric-pericentromeric region with the organization of the human counterpart at 6q24.3 might allow us to examine hypotheses of the formation and progression of ENCs.

Human BAC RP11-474A9 (L2 in Fig. 2) mapping at chr6:145,651,644 to 145,845,896 yielded an apparently splitting signal around the MMU4 centromere (*9*). It was therefore considered to be the probable seeding point, and the FISH analysis of flanking markers was consistent with this conclusion (Fig. 3). The construction of a BAC contig spanning the MMU4 centromere started, therefore, from the L2 locus and is reported in detail in the supporting online material (SOM) text. To briefly summarize this construction, appropriate human sequence tagged sites (STSs), mapping within 1 megabase from both sides of the MMU4 centromere, were used to screen high-density filters of the macaque BAC library CH250, segment 1 (http://bacpac.chori.org). FISH analysis of these BACs showed that some of them were duplicated on both sides of the MMU4 centromere. The sequencing of BAC ends and of appropriate polymerase chain reaction (PCR) products and FISH
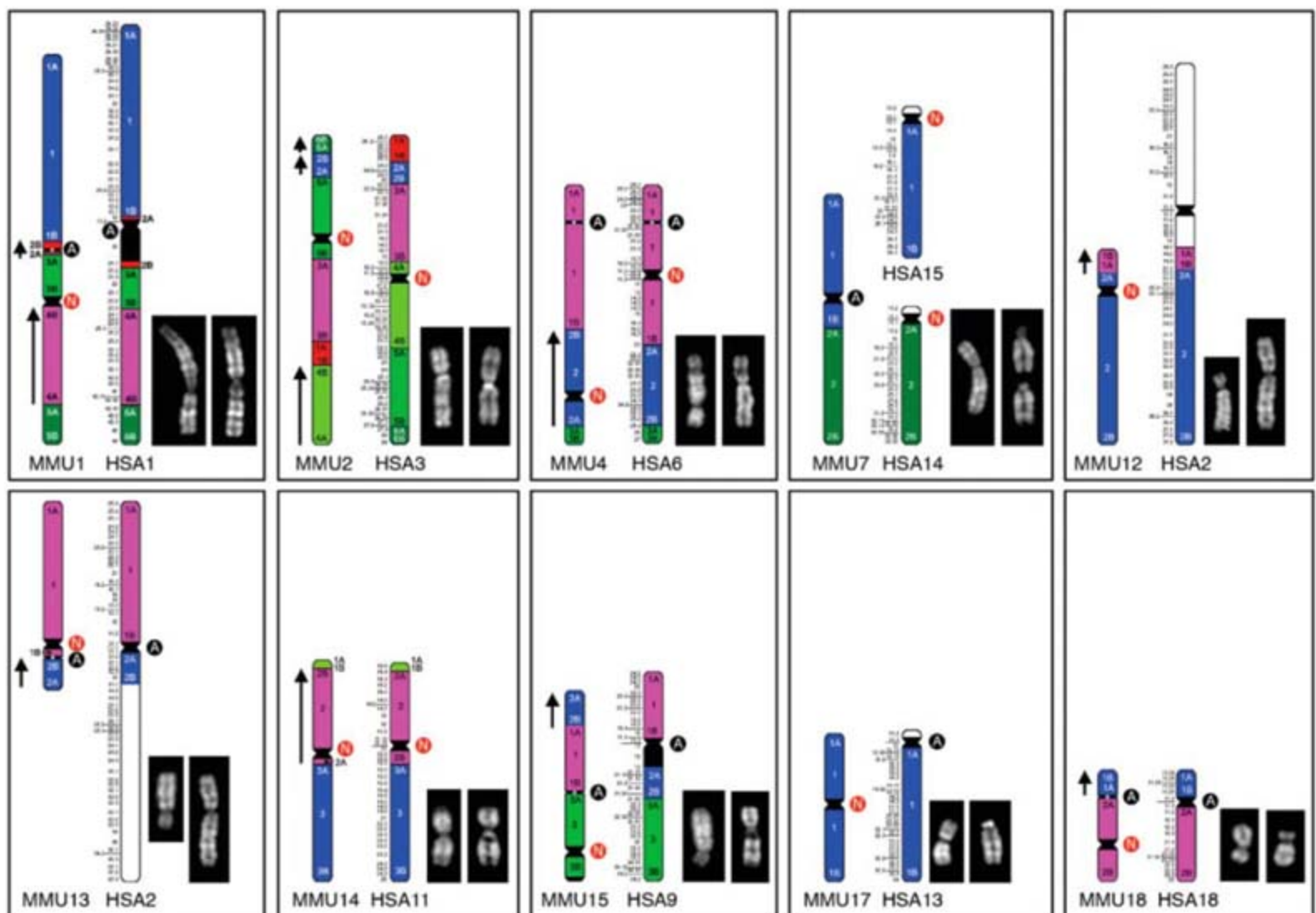
Fig. 1. All macaque (left) or human (right) chromosomes harboring ENCs. Quinacrine mustard banded macaque (left) and human (right) chromosomes are also displayed. Details are reported at www.biologia.uniba.it/macaque. Some macaque chromosomes are shown upside down to facilitate comparison in accordance with the orientation in the whole-genome assemblage. The Ns in red circles flank ENCs; the As in black circles flank ancestral centromeres. Synteny block conservation is indicated by the color. The arrows on the left of some macaque synteny blocks indicate that these blocks have opposite sequence polarity, with respect to the corresponding block in humans. These annotations, obviously, are affected by the upside down position of some chromosomes. The Human Genome Project has assumed as "+" sequence polarity the 5'->3' strand that starts from the tip of the short arm of each chromosome. The assumption that centromeres were conserved leads to a violation of this rule for macaque chromosomes 1 (1), 2 (3), 4 (6), 10 (20/22), 13 (2p), and 18 (18) (corresponding human chromosomes in parentheses).

experiments on stretched chromosomes (Fig. 3D) allowed for the construction of a contig defining the duplicated pericentromeric region. In summary, seven imperfect copies of a 250-kb segment, mapping at the seeding point, were duplicated both proximally and distally to the MMU4 centromere. The global tiling path and the detailed organization of the central duplicated region are shown (Fig. 4, A and B). Duplicated regions appear to be co-oriented with respect to each other and to the human sequence assembly. Unexpectedly, a high proportion of BAC ends of duplicated clones were alphoid in nature (30 out of 46). These alphoid sequences showed a monomeric structure that is typical of peripherally located alpha satellite sequences (table S8). Small, repeated inversions in the duplicated regions might be hypothesized to account for these findings. This hypothesis, however, clashes with the apparent co-orientation of the duplicated blocks.

Data from human pericentromeric regions have shown that the ratio of inter- versus intra-chromosomal duplications is about 6:1 (*1*). We previously suggested that duplications of the ENC of macaque chromosome 17 (human 13) were intrachromosomal only (*3*), but in that case only human probes were used, which precluded any firm conclusion on the absence of interchromosomal duplications. Our current results show that the MMU4 pericentromeric duplications detected by FISH were strictly intrachromosomal and originated only from the ENC seeding point. We have also shown that centromeres of human chromosomes 3, 6, 11, 14, and 15 are ENCs (*6, 9–11*). Chromosomes HSA3 and HSA6 match the pattern we found on MMU4, whereas human chromo-somes 11, 14, and 15 accommodate large blocks of interchromosomal duplications (*1*). A careful analysis of the evolutionary history of the latter chromosomes, however, showed that it was very likely that large blocks of segmental duplications were already present or simultaneously seeded in the ENC region (*10, 11*). It could therefore be hypothesized that a novel centromere triggers only local duplication activity, whereas interchromosom-al duplications are triggered by distinct forces, probably linked to intrinsic properties of specific sequences (*17, 18*). However, until further cases are studied we cannot rule out that the duplications we detected on MMU4 are simply macaque-specific.

The corresponding human region in proximity to the L2 marker was investigated for gene content. A relatively large region (780 kb) harboring the MMU4 centromere has not been annotated in the UniProt, RefSeq, and GenBank mRNA databases. The two closest genes on opposite sides, UTRN (chr6:144,654,658 to 145,209,657) and EPM2A (chr6:145,988,133 to 146,098,299), are 778 kb apart. Heterochromatin supposedly silences embed-ded genes (*19*). Genes mapping to regions where a centromere repositioning occurred might be at risk of silencing, but recent reports have indicated that a neocentromere by itself does not repress gene expression (*20–22*). The gene silencing might be attributed to the successive heterochromatization of the region. The average gene content in the human genome is about 1 gene per 100 kb (www.ncbi.nlm.nih.gov). Human chromosome 6 contains about 1272 genes, on average 1 gene every 131 kb (*23*). Consequently, six genes would be expected in the 778-kb gene-desert area. A similar gene-desert area was also found around the ENC of the Old World monkey chromosome homologous to hu-man chromosome 13 (*3*). Our data appear to support the hypothesis that the absence of genes in the ENC seeding region can play an important role in ENC maintenance and progression. Analyses of additional ENCs and their corresponding regions in the human genome will be required to determine whether this is a stochastic occurrence or whether it represents a prerequisite for novel centromere survival. This hypothesis initially appears to be contradicted by the presence of active genes at the centromeres of rice chromosomes 8 (*24*) and 3 (*25*). However, Nagaki *et al.* and Yan *et al.* suggest that these two rice centromeres may represent ENCs that are still acquiring the full heterochromatic organization that is typical of normal centromeres, and the analysis of the fully sequenced *Arabidopsis* genome strongly supports the view that the absence of gene expression in centromeres is also a general rule in plants. Alternatively, it could be hypothe-sized that the heterochromatization process pushes the surrounding genes to pericentromic regions without affecting their expression.

Ferreri *et al.* (*26*) reviewed the various hypotheses formulated to explain ENC and clinical-neocentromere emergence. One hypothesis proposes that the centromere seeding event is essentially epigenetic in nature and is sequence independent (*27*). Another hypothesis considers the seeding regions to be domains with inherent latent centromere-forming potentiality (*11, 28*). A third hypothesis suggests that rearrangements trigger neocentromere seeding through chromatin repatterning (*11*). Roizes (*29*) has suggested that damage to a centromere, like retroposon insertion, could trigger the emergence of evolutionary neocentro-meres. All of these hypotheses consider clinical neocentromeres and ENCs to be strictly related.

An unexpected finding is the high number of ENCs in recent human and Old World monkey evolution. In the 25 million years since macaque and human divergence, 14 ENCs have arisen and become fixed in either the human or the macaque lineage. It is difficult to escape the conclusion that ENCs had a considerable impact on shaping the primate genome and that they are fundamental to our understanding of genome evolution. Knowl-edge of centromere repositioning, for instance, provides a cogent explanation for the unusual clus-tering of human clinical neocentromeres at 15q25, the domain of an inactivated ancestral centromere (*11*). Despite their relevance, ENCs have never been identified on the basis of sequence analysis alone. Indeed, the extensive pericentromeric dup-lication we report has not been identified in the macaque genome assembly, reinforcing the opinion
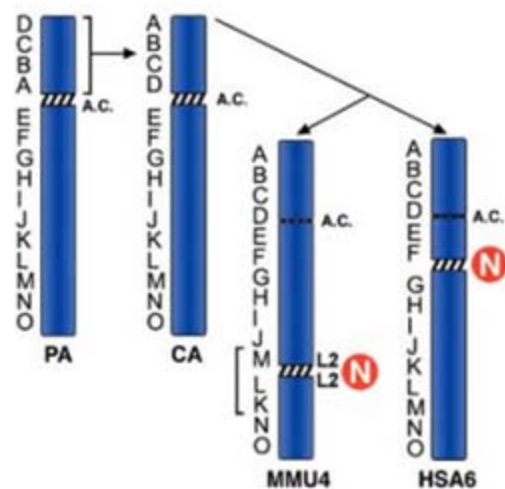

Fig. 2. A reconstruction of chromosome 6 evolution in primates [modified from (*9*)]. The letters indicate the specific BAC clones used in the study (*9*) and reported in table S9. PA, primate ancestral; CA, Catarrhini ancestral; and A.C., ancestral centromere. The letter N in a red circle represents a novel centromere. A square bracket encompasses the Old World monkey–specific (macaque included) inversion of K-L-M markers, with respect to the human form. MMU4 is upside down, with respect to the correct position as reported in Fig. 3, to allow for an easy synteny comparison. L2, human BAC RP11-474A9.

Fig. 3. (**A** to **C**) Partial metaphases showing examples of FISH experiments on macaque and human chromosomes, with the use of the non-duplicated BAC clones (A) CH250-209i5 and (B) CH250-215i15 and the duplicated clone (C) CH250-284C24. (Mapping details are given in Fig. 4, A and B.) (**D**) Partial metaphases showing FISH on stretched chromosomes with the use of the duplicated BAC clones CH250-188G18 (red) and CH250-67B18 (green), containing alphoid sequences. The arrows indicate MMU4.
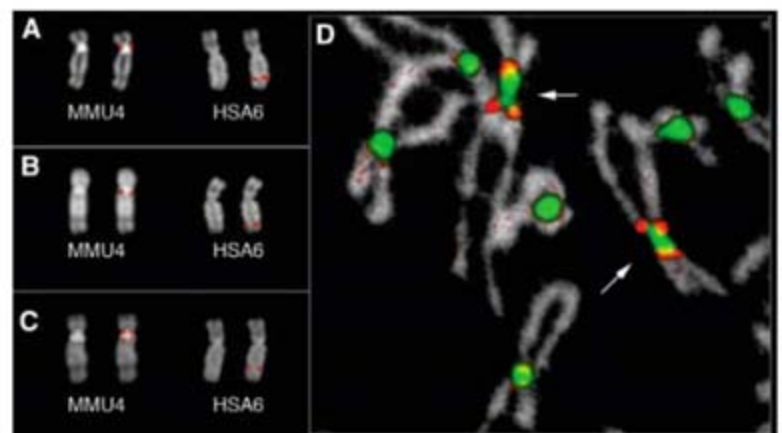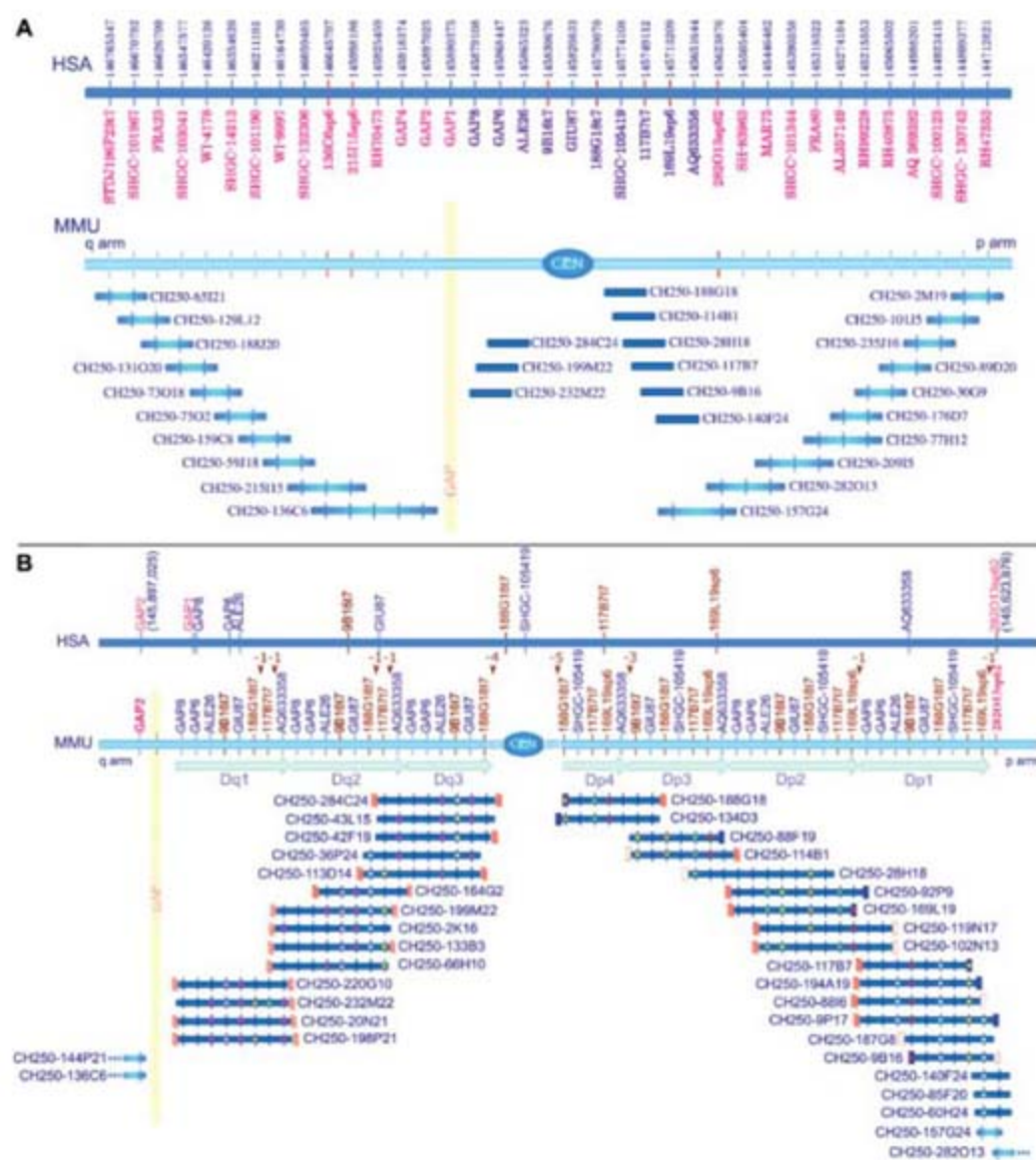
The Rhesus Macaque Genome

Fig. 4. (**A**) BAC clone contig map of the MMU4 region corresponding to the HSA6 DNA segment from bp 144,712,821 (right) to 146,765,547 (left) (UCSC March 2006 release). Human and macaque STSs (table S4) were used to screen high-density filters of the macaque BAC library CH250. Positively hybridizing BACs were then used in FISH experiments on macaque and human metaphases, and a contig tiling path of the BACs (blue segments) was defined. The analysis discovered single and duplicated regions around the MMU4 pericentromeric region. The marker positions of the STSs in the human sequence are reported in the top line. STSs in red are localized in the MMU4 single-copy region; blue STSs markers are within the region that is duplicated around the MMU4 centromere. This duplicated region is illustrated in detail in (B). Vertical red lines represent macaque STSs derived from BAC-end sequences. The yellow vertical region represents a gap that is probably composed of sequences refractory to cloning. (**B**) Details of the BAC contig assembly of the duplicated pericentromeric region of MMU4. FISH analysis of the macaque BAC clones indicated that duplicated blocks were present on both sides of the centromere. To further characterize these duplicated pericentromeric regions, we sequenced and aligned appropriate STS-amplified products from different clones with the MegAlign software (31). The sequence analysis allowed for the classification of the duplicated STSs and the corresponding BACs into seven separate blocks: three on the q arm (Dq1-3) and four on the p arm (Dp1-4) (table S7). The top blue line shows the mapping of the STS markers in human. The first nonduplicated STS, on both sides, is shown in red. The MMU line depicts all of the STS that were used in the study, arranged according to their inferred position in macaque. Macaque STSs derived from BAC-end sequences are shown in brown. Dq1, Dq2, and Dq3 and Dp1, Dp2, Dp3, and Dp4 indicate the duplicated subsets on the q-arm side (Dq-block) and on the p-arm side (Dp-block) of the centromere, respectively. Some STSs are apparently missing in some blocks. Their positions are indicated by brown arrowheads above the macaque STS line. The number of missing STSs is also indicated. Sequenced BAC ends are represented by small rectangles at the borders of the BACs. White and red diagonally striped rectangles stand for monomeric alpha satellite sequences; unfilled red rectangles stand for repetitive nonsatellite elements; and blue rectangles stand for single-copy sequences. The absence of a rectangle indicates that the end sequencing failed twice. Sequenced PCR products of specific STS primers are represented by colored circles. These sequences were used for the classification of duplicated blocks. The colored circles indicate STSs whose PCR products have been sequenced. Circles related to the same STS are in the same color only if their sequences are perfectly matched. A full description of the BAC contig assembly is reported in the SOM text.

that an integrated, multidisciplinary approach is needed for high-quality genome assembly and for comparative genomics (30).

The present data extend the link between segmental-duplication bias and centromeres to additional primate species. The homology and shuffling of sequences creates substrates for evolutionary innovation (the birth of new genes) and instability (via non-allelic homologous recombination). Lastly, the contig assembly we have constructed represents a framework for the complete sequencing of the pericentromeric region of MMU4 ENC through a direct sequencing of BAC templates, as opposed to whole-genome shotgun sequencing.

References and Notes

1. X. She et al., Nature 430, 857 (2004).
2. G. Montefalcone, S. Tempesta, M. Rocchi, N. Archidiacono, Genome Res. 9, 1184 (1999).
3. M. F. Cardone et al., Genome Biol. 7, R91 (2006).
4. D. J. Amor, K. H. Choo, Am. J. Hum. Genet. 71, 695 (2002).
5. D. J. Amor et al., Proc. Natl. Acad. Sci. U.S.A. 101, 6542 (2004).
6. M. Ventura et al., Genome Res. 14, 1696 (2004).
7. W. J. Murphy, L. Fronicke, S. J. O'Brien, R. Stanyon, Genome Res. 13, 1880 (2003).
8. R. Marzella et al., Genomics 63, 307 (2000).
9. V. Eder et al., Mol. Biol. Evol. 20, 1506 (2003).
10. M. F. Cardone et al., Genomics, in press.
11. M. Ventura et al., Genome Res. 13, 2059 (2003).
12. J. Wienberg, R. Stanyon, A. Jauch, T. Cremer, Chromosoma 101, 265 (1992).
13. S. Muller, J. Wienberg, Hum. Genet. 109, 85 (2001).
14. W. J. Murphy, R. Stanyon, S. J. O'Brien, Genome Biol. 2, REVIEWS0005 (2001).
15. R. Stanyon et al., Am. J. Primatol. 50, 95 (2000).
16. X. She et al., Genome Res. 16, 576 (2006).
17. J. E. Horvath et al., Genome Res. 15, 914 (2005).
18. J. A. Bailey, E. E. Eichler, Nat. Rev. Genet. 7, 552 (2006).
19. K. S. Weiler, B. T. Wakimoto, Annu. Rev. Genet. 29, 577 (1995).
20. R. Saffery et al., Mol. Cell 12, 509 (2003).
21. N. C. Wong et al., PLoS Genet. 2, e17 (2006).
22. A. L. Lam, C. D. Boivin, C. F. Bonney, M. K. Rudd, B. A. Sullivan, Proc. Natl. Acad. Sci. U.S.A. 103, 4186 (2006).
23. A. J. Mungall et al., Nature 425, 805 (2003).
24. K. Nagaki et al., Nat. Genet. 36, 138 (2004).
25. H. Yan et al., Plant Cell 18, 2123 (2006).
26. G. C. Ferreri, D. M. Liscinsky, J. A. Mack, M. D. Eldridge, R. J. O'Neill, J. Hered. 96, 217 (2005).
27. A. Alonso et al., Hum. Mol. Genet. 12, 2711 (2003).
28. K. H. A. Choo, Am. J. Hum. Genet. 61, 1225 (1997).
29. G. Roizes, Nucleic Acids Res. 34, 1912 (2006).
30. M. Rocchi, N. Archidiacono, R. Stanyon, Genome Res. 16, 1441 (2006).
31. Materials and methods are available as supporting material on Science Online.
32. We acknowledge the Ministero della Universita' e della Ricerca (MUR) and the European Commission (grant QLRI-CT-2002-01325) for financial support. This work was also supported in part by NIH grants GM58815 and HG002385 to E.E.E. E.E.E. is an investigator of the Howard Hughes Medical Institute. We would like to thank G. Herrick for critical reading. R.S. was supported by the MUR grant "Mobility of Italian and Foreign Researchers Residing Abroad."

The Macaque Genome

CONTRIBUTORS

Richard A. Gibbs, R. Alan Harris, and Aleksandar Milosavljevic, Baylor College of Medicine; Judy Cameron and Vince Warren, Oregon Health and Science U; Ashley Haase and Qingsheng Li, U Minnesota; Miles Herkenham and David Leopold, NIH/NIMH; Steve Suomi, NIH/NICHD; Arthur M. Lesk and Webb Miller, Pennsylvania State U; Brian Raney, UC Santa Cruz; Adam Siepel and Tomas Vinar, Cornell U; and the Rhesus Macaque Genome Sequencing and Analysis Consortium

COORDINATORS
Barbara R. Jasny, Laura M. Zahn

DESIGN & ART DIRECTION
Kelly Krause

ILLUSTRATION
Katharine Sutliff

PRODUCTION
Marcus Spiegler
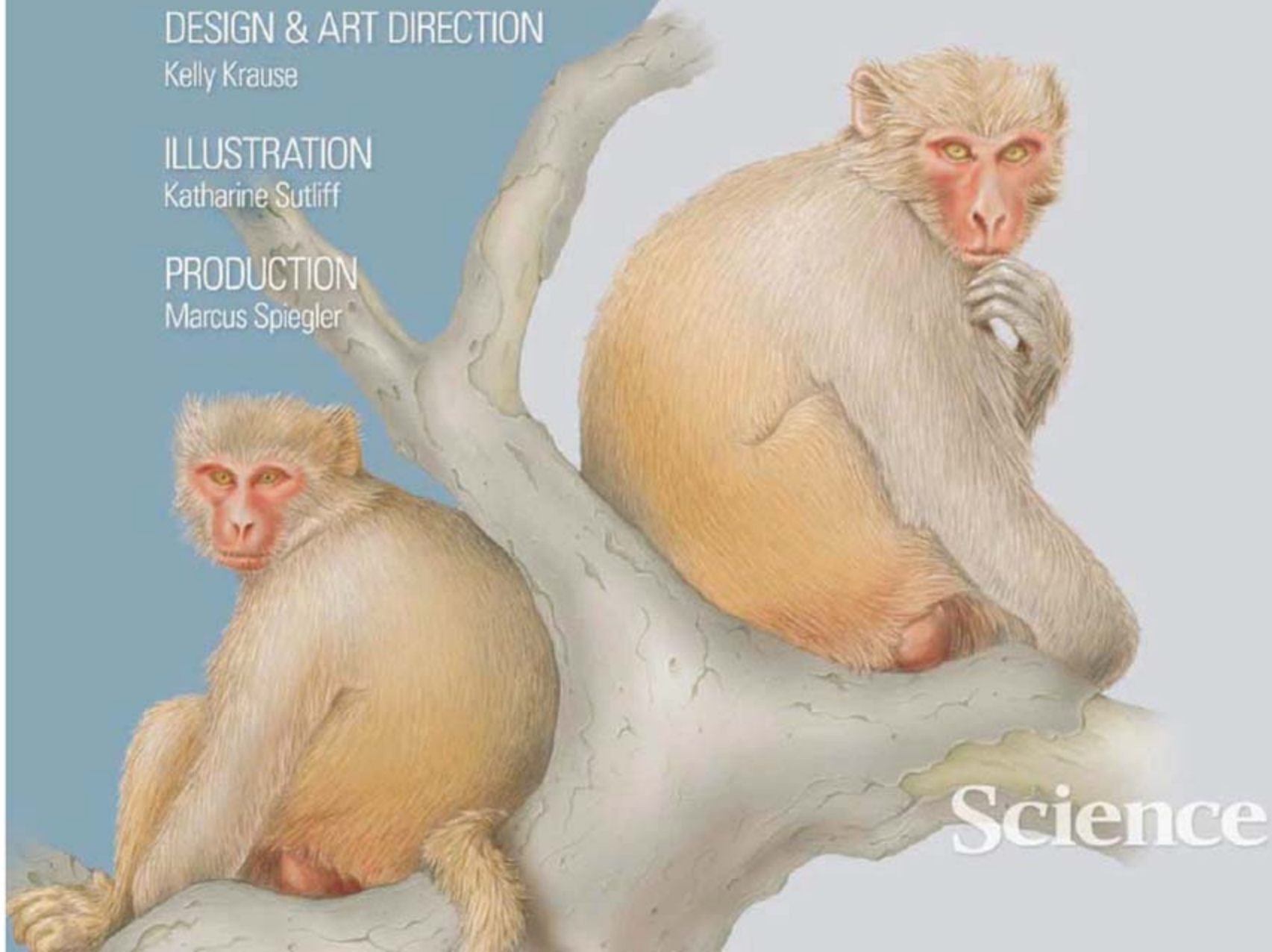
Science

The Macaque Genom[e]

Evolution

THE RHESUS MACAQUE (*Macaca mulatta*) facilitates an extraordinary range of biomedical and basic research, and the publication of the genome only makes it a more powerful model for studies of human disease. Despite distinct differences, the macaque genome mirrors our own in complexity and is 93% similar to humans. Because of the macaque's position relative to humans and chimpanzees, we now have a unique opportunity to learn about the diverse evolutionary processes that we have undergone since our primate ancestors diverged 25 million years ago (Ma). Additional material is available online at www.sciencemag.org/sciext/macaqueposter/.

Positive Selection Across the Genome

NATURAL SELECTION of genetic variants associated with advantageous traits plays an important role in the course of evolution. Evidence of the evolutionary process can be found by comparing the DNA sequences of related species. The cytogenetic map below indicates the approximate locations of the 67 genes showing the strongest evidence for positive selection in primates, on the basis of a comparative analysis of the macaque, human, and chimpanzee genomes. Colors denote Gene Ontology (GO) categories. *Image: R. Alan Harris and Aleksandar Milosavljevic, Baylor College of Medicine; Tomas Vinar, Cornell University*

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

X

- Immunity and defense
- Signal transduction
- Cell adhesion
- Catalytic activity
- Keratin and keratin-associated
- Miscellaneous and unknown

Gene Family Expansions

A PHYLOGENETIC TREE of PRAME genes (a cancer-associated gene family normally expressed in the testes). Estimates of copy number suggest this family expanded dramatically prior to the evolution of Old World monkeys. More recently, PRAME genes underwent independent expansions in humans and chimpanzees while remaining relatively stable in the macaque lineage. This illustrates how the macaque genome helps unravel complex evolutionary histories. *Image: Brian Raney, UC Santa Cruz; Tomas Vinar and Adam Siepel, Cornell University*

25 Ma

4–8 Ma

Macaca mulatta 8 genes

Pan troglodytes 12 genes

Homo sapiens 26 genes

Mouse
Rat
Dog
Chimpanzee
Human
Macaque
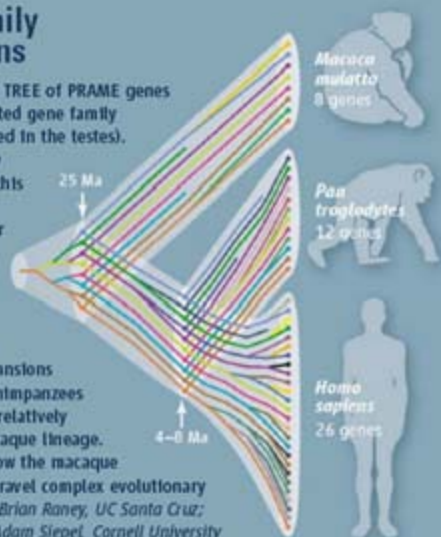
Gene Evolution

THE UPPER PANEL SHOWS the ancestral structure of the TSPAN8 gene, reflected in the mouse, rat, dog, chimpanzee, and human genomes, and the presence of an additional exon in the macaque genome. A macaque-specific segmental duplication created two copies of an exon, which now has mutually exclusive alternatively spliced forms. In addition, these exons appear to have experienced strong positive selection since the duplication. The lower panel shows that in one of these exons 9 of 44 amino acids differ between humans and macaques. *Image: Adam Siepel, Cornell University*

Science

Neurobiology

THE IMAGES ABOVE SHOW only two of the myriad ways in which the macaque brain has already been mapped. Now, genomic information can be overlaid on these maps—e.g. by providing high-resolution data on gene expression or in creating specific probes and markers—to provide insight into neurodevelopment, behavior, and neurological disease. Images: Left: Intersecting planes centered on a gadolinium injection site in the thalamus. *Image: David Leopold, NIH/NIMH* Right: A colorized brain section showing opiate receptor binding in the hypothalamus and amygdala. *Image: Miles Herkenham, NIH/NIMH*

Disease-Related Alleles

ONE OF THE SURPRISIN findings from the macaque genome is that the form of several genes that are associated with disease in humans is the norm in healthy macaques. The figure at right is of one subunit of the tetrameric human phenylalanine hydroxylase, with side chains shown for positions at which the sequences of human and macaque differ. The side chain shown in cyan is for tyrosine in the normal human sequence. A mutation to histidine in the human is associated with phenylketonuria, but corresponds to the wild-type residue at this position in the macaque. *Image: Arthur M. Lesk, Pennsylvania State University, based on Protein Data Bank entry 2PAH*
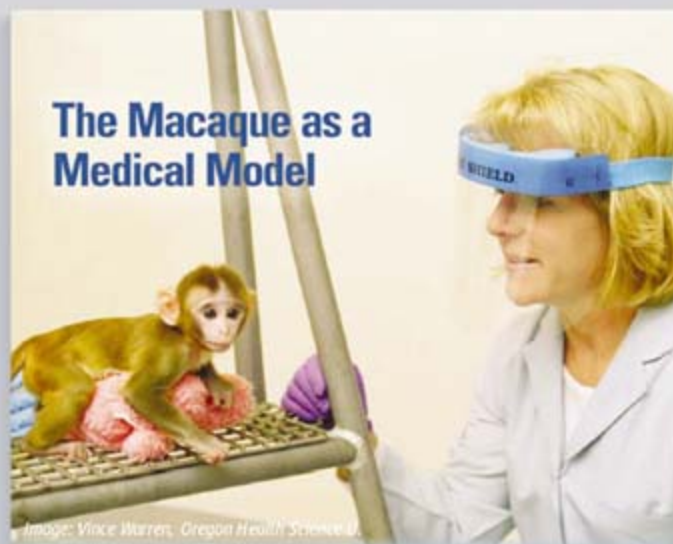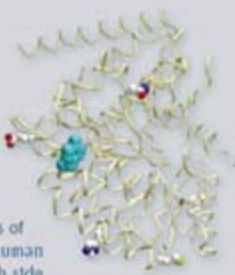
The Macaque as a Medical Model

Image: Vince Warren, Oregon Health Science U.

Behavior
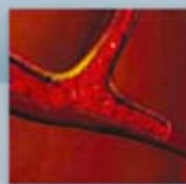
<< Cardiovascular Disease

MALE: GETTY

Aging >>

Diabetes

Cancer

Drug Abuse
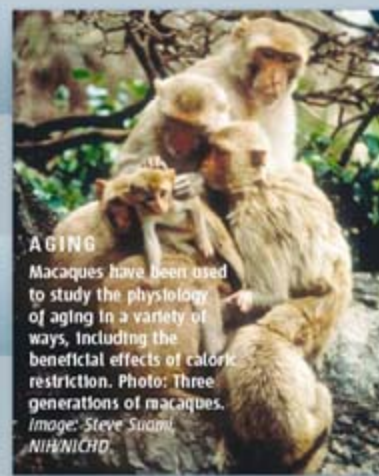
AGING
Macaques have been used to study the physiology of aging in a variety of ways, including the beneficial effects of caloric restriction. Photo: Three generations of macaques. *Image: Steve Suomi, NIH/NICHD.*

Diet and Health

Endocrinology

HIV/AIDS
Macaque cervical tissues 7 days after vaginal mucosal exposure to simian immunodeficiency virus (SIV). Bright cells are SIV-infected. *Images Ashley Haase and Qingsheng Li, University of Minnesota*

<< Infectious Diseases

Reproductive Physiology

Osteoporosis >>

IMAGE: GETTY

Biomedicine

A Symmetric Bipolar Nebula Around MWC 922

P. G. Tuthill[1]* and J. P. Lloyd[2]

The Be star MWC 922 has previously been noted for its spectral features: unusually strong forbidden [Fe II] in emission (*1*) and more recently the dust mineralogy and chemistry of its rich infrared spectrum (*2*). The distance to this object is unknown (it may lie within the Ser OB1 association at 1.7 kpc), as is its evolutionary status (both pre- and postmain sequence have been suggested).

Here, we report high angular resolution observations in the infrared H band made with the adaptive optics system on the 200" Palomar telescope [detailed further in (*3*)]. An image showing the extended Red Square nebula surrounding MWC 922 is given in Fig. 1, together with a highpass filtered image that reveals underlying structural elements and a model describing the key features. The box-shaped X structure necks down into twin opposing hyperbolic bicones separated by 0″.34 at the center, where they are crossed by an equatorial dark band running northeast to southwest. Along the principle axis of symmetry (PA = 46°), a remarkable series of orthogonal linear rungs appear at 1″.77 and 3″.1 arc sec (northwest bicone) and at –1″.44 and –3″.2 arc sec (southeast bicone). Where the rungs meet the bicone surfaces, we find bright vortices that subtend an opening angle of 105° at the origin.

In addition, we have established the presence of a series of radial linear features resembling a comb and appearing to lie along both bicone surfaces between the second bright rung and the outer edge of the nebula. One possibility is that the comb may be the outcome of a projected illumination effect as light from the central regions is blocked by some periodic structure en route to the outer bicones. An example could be shadows cast by "ripples" on the rim of a circumstellar disk, such as those believed to result from gravitational or other excitation of standing-wave modes (*4*), although high azimuthal orders ($l \sim 50$) are required.

Among the most notable properties of this nebula is the extent to which structures are reflection-symmetric about the principle axis. This symmetry and linearity of the rungs implies that the viewing angle onto this system axis is very close to 90°. It seems likely that this appearance depends on this critical alignment, whereas the regularity of form argues against a premain sequence identity. A weaker (although still notable) symmetry exists across the equator between structures in the northwest and south-

east bicones, with perhaps the largest departure being the axial displacement of the innermost rung, which appears 20% further out in the
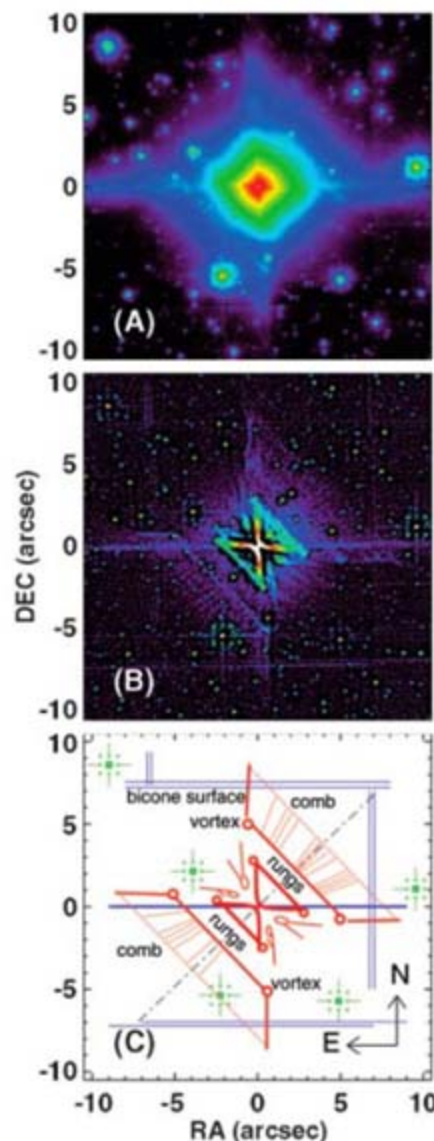
Fig. 1. MWC 922 from Palomar H-band adaptive optics imaging: (**A**) image at a logarithmic color stretch, (**B**) image data after Laplacian filtering to accentuate structure [such as edges; see (*3*) for details], and (**C**) model skeleton structural elements fitted to the Laplacian filtered image. This model also depicts spurious linear features from imperfect mosaicing and charge persistence (blue) and bright neighboring stars (green), whereas real Red Square nebula structures are plotted in red. The system's principal symmetry axis is given as a dot-dashed line, together with a compass rose and annotations labeling key features in black. DEC, declination; RA, right ascension.

northwest than the southeast bicone. This high degree of correspondence between structures in the two bicones argues for common creation in episodic eruptions within the central star system. This result also has implications for other bipolar systems, such as the Red Rectangle, where direct correspondence between pairs of rungs in the two bicones could not be established (*5*), arguing against models with one-sided mass ejections.

Adopting the (speculative) distance of 1.7 kpc, the linear size of the Red Square is twice as large as the Red Rectangle at the same dynamic range in surface brightness. If we further speculate that outflow velocities are not dissimilar [~7 km/s (*5*)], then the dynamical ages of the two major rung systems are ~3000 and 6000 years. However, given the significantly hotter spectrum of MWC 922's central star [B3-B6 (*6*)] as compared with the Red Rectangle [peculiar A supergiant in a 319-day binary (*5*)], basic parameters such as wind speeds may differ substantially. The finding of a cousin to the Red Rectangle, but with differing spectrum and a hotter central star, implies that the conditions of formation for these elegant bipolar ladders may not be so singular and unique as formerly thought.

The clearly demarked narrow rungs reported here places MWC 922 in rare company with only two other known systems [Sher 25 and HD168625 (*7*)] that support ringed bipolar nebulae, structures which make them viable progenitors for polar ring systems like that seen in SN1987A. As the only one of the three to exhibit multiple rings and a wealth of other forms, MWC 922 should make an excellent laboratory for the study of the creation of such structures.

References and Notes

1. R. J. Rudy, P. Erwin, G. S. Rossano, R. C. Puetter, *Astrophys. J.* **398**, 278 (1992).
2. F. J. Molster, L. B. F. M. Waters, A. G. G. M. Tielens, M. J. Barlow, *Astron. Astrophys.* **382**, 184 (2002).
3. Materials and methods are available on *Science* Online.
4. S. H. Lubow, J. E. Pringle, *Astrophys. J.* **409**, 360 (1993).
5. M. Cohen, H. Van Winckel, H. E. Bond, T. R. Gull *Astron. J.* **127**, 2362 (2004).
6. M. Jura, J. Turner, S. P. Balm, *Astrophys. J.* **474**, 741 (1997).
7. N. Smith, *Astron. J.*, in press; available at http://arxiv.org/abs/astro-ph/0611544.
8. We thank Palomar Observatory and its dedicated staff for help with the observations.

[1]Physics Department, University of Sydney, NSW 2006, Australia. [2]Department of Astronomy, Cornell University, Ithaca, NY 14853, USA.

*To whom correspondence should be addressed. E-mail: p.tuthill@physics.usyd.edu.au

Quantum Structure of the Intermolecular Proton Bond

J. R. Roscioli, L. R. McCunn, M. A. Johnson*

A proton shared between two closed-shell molecules, [A·H⁺·B], constitutes a ubiquitous soft binding motif in biological processes. The vibrational transitions associated with the shared proton, which provide a direct probe of this interaction, have been extensively studied in the condensed phase but have yielded only limited detailed information because of their diffuse character. We exploited recent advances in gas-phase ion spectroscopy to identify sharp spectral features that can be assigned to both the shared proton and the two tethered molecules in a survey of 18 cold, isolated [A·H⁺·B] ions. These data yield a picture of the intermolecular proton bond at a microscopic scale, facilitating analysis of its properties within the context of a floppy polyatomic molecule.

A general property of Lewis bases is their ability to efficiently trap excess protons on their exposed electron lone pairs to form covalent hydrides. When two such molecules (A and B) compete for the same proton, the resulting binary complex, [A·H⁺·B], becomes tightly bound in a cationic analog of the familiar H bond, providing a "point-contact" coupling between lone pairs on electronegative atoms (such as N and O) embedded in larger molecular scaffolds. Practically, these intermediate-strength couplings play an important role in biological and physical processes requiring proton transport in hydrophobic media (such as transmembrane proton pumps) (1–4). From the physical point of view, the motif involves trapping a subatomic particle in a soft, three-dimensional potential well, the shape of which is determined by the chemical properties of the flanking atoms. Not surprisingly, theoretical treatment of this interaction is complicated because the low mass of the proton results in it undergoing large-amplitude motion (as large as 0.3 Å in the complexes we studied), even in its vibrational zero-point level (5–9).

These binary systems are of particular interest because they are so stable that the symmetrical complexes, such as [A·H⁺·A], can be prepared in macroscopic quantities by careful stoichiometric control of salt solutions (10, 11). Over the past decade, this situation has been exploited to survey the spectral signatures associated with the shared proton in room-temperature condensed phase samples, for which the bands attributed to the shared proton generally appear as diffuse absorptions in the 1000 cm⁻¹ range (10, 11). One notable finding in the recent condensed phase studies involves a curious disappearance of the vibrational transitions nominally associated with the flanking organic groups in symmetrical [A·H⁺·A] complexes (10, 11).

Rationalizations of this behavior include broadening of the A molecule's transitions into a continuum background, possibly as a result of their modulation by the stochastic motion of the bridging proton. Unfortunately, the large breadths of the spectral envelopes recovered in solution, which are characterized by Gaussian contributions with widths of 200 to 400 cm⁻¹, encompass nearly the entire range of the expected intramolecular (e.g., C-O and P-O stretch) transitions. Thus, it has been difficult or impossible to quantify exactly how these intramolecular bands evolve upon complexation or to establish the intrinsic pattern of homogeneous bands present when fluctuations are frozen out at 0 K.

Because such unresolved spectra mask the inherent molecular eigenstates at play, there has been a parallel effort by gas-phase ion spectroscopists to study these complexes under isolated conditions. A major advantage of studying these systems in isolation is that the resulting discrete spectra should be amenable to analysis with the full arsenal of tools developed through decades of work on the electronic structure and vibrational dynamics of polyatomic molecules. There have been recent reports on spectra of the isolated gas-phase complexes [e.g., (Me₂O)₂H⁺, where Me = CH₃], obtained using infrared multiphoton dissociation (IRMPD) (12–14). The resulting spectra consisted of a diffuse system of bands with comparable intensity throughout the low-energy region of 700 to 1500 cm⁻¹, not unlike those observed in the condensed phase (11). The similarity of these broad structures for several gas-phase complexes led Moore et al. to suggest that this extended suite of bands together comprised the signature of the O-H-O bridge. Because both the solution and IRMPD studies were carried out at room temperature, however, it is not clear to what extent this broadening results from thermal excitation as opposed to the intrinsic [A·H⁺·B] spectral character.

Revealing inherent linewidth through argon nanomatrices. The goal of our study was to overcome the complications introduced by the inherent nonlinearity in IRMPD and the thermal effects present at elevated temperatures by acquiring spectra of the vibrationally cold, gas-phase [A·H⁺·B] complexes with the use of rare-gas nanomatrix spectroscopy (15). We focused on the proton-bound complexes involving methanol (MeOH), ethanol (EtOH, where Et = CH₂CH₃), and ethers (Me₂O and Et₂O), in addition to the small molecules H₂O, NH₃, CO₂, and the rare gases (RG) Ar and Xe. To aid in the assignment of the resulting band patterns, we also obtained the spectra of the protonated monomers (e.g., Me₂OH⁺) which, when combined with the known spectra of the neutral species (e.g., Me₂O), allowed us to learn how the constituent molecules in the [A·H⁺·B] complex flex and evolve as the proton becomes preferentially attached to the stronger base.

Vibrational spectra of [A·H⁺·B]·Ar complexes were acquired over the range of 600 to 4000 cm⁻¹ in a linear action mode with the use of the Yale double-focusing, tandem time-of-flight photofragmentation spectrometer (16). In most cases, this instrument synthesizes the proton-bound binary complexes in an ionized supersonic free-jet expansion that contains trace water vapor seeded in Ar. Typically, this first generates H₃O⁺·Arₙ cluster ions, which are then interfaced with organic reagents that are entrained into the expansion on the low-pressure side through two independently controlled pulsed valves. The likely sequence of events

$$H_3O^+\cdot Ar_n + A \rightarrow H_3O^+\cdot A\cdot Ar_m + (n-m)\,Ar \quad (1)$$

$$H_3O^+\cdot A\cdot Ar_m + B \rightarrow$$
$$[A\cdot H^+\cdot B]\cdot Ar + H_2O + (m-1)\,Ar \quad (2)$$
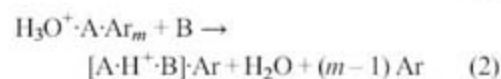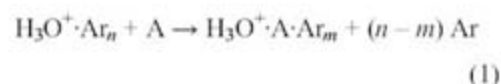
creates the [A·H⁺·B]·Ar target ions. Some cases, such as MeOH·H⁺·CO₂, required variation of this scheme in which MeOH was entrained into the Ar jet to form the MeOH·H⁺·Arₙ clusters before the introduction of CO₂. Formation of these rare-gas complexes is crucial, because they are sufficiently weakly bound to enable detection of their low-energy vibrational spectra by very sensitive "messenger atom" action spectroscopy (17):

$$[A\cdot H^+\cdot B]\cdot Ar + h\nu \rightarrow [A\cdot H^+\cdot B] + Ar \quad (3)$$
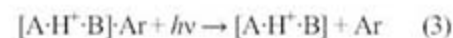
Because this process is carried out in a one-photon regime, it circumvents the complications inherent in the IRMPD approach, for which as many as 16 photons are required to dissociate these very strongly bound complexes (binding energy $D_0 \sim$ 1.37 eV in H₅O₂⁺) (18). In addition, the weakly bound Ar serves to limit the internal energy content so that the resulting spectra largely reflect the properties of the vibrational zero-point level. Selective detection of the lighter [A·H⁺·B] photofragment ions (Eq. 3) after initial isolation of the heavier [A·H⁺·B]·Ar complex yields a

Sterling Chemistry Laboratory, Yale University, Post Office Box 208107, New Haven, CT 06520, USA.

*To whom correspondence should be addressed. E-mail: mark.johnson@yale.edu

background-free scheme to monitor the absorption of infrared photons down to 600 cm^{-1} (19). Spectra are displayed as the net fragmentation signal normalized to the laser output energy per shot at each excitation wavelength.

Probing the asymmetric stretching vibrational transition (ν_{sp}). At the simplest level, we expected three vibrational fundamentals of the shared proton, reflecting its three-dimensional confinement: the asymmetric stretch parallel to the heavy atom axis [$\nu_{sp}(\parallel)$] and two transitions arising from vibrations roughly perpendicular to this axis [$\nu_{sp}(\perp)$]. Even at the harmonic level, however, shared-proton displacements participate in many of the calculated normal modes describing the collective fundamental motions in the complexes. The infrared intensity of the mode containing mostly ν_{sp} ($\parallel$) character is calculated to dominate those arising from more perpendicular motions. The $\nu_{sp}(\parallel)$-based transitions should therefore provide the clearest probe of the intermolecular proton bond. They are expected to be both intense and quite responsive to the details of the bond, spanning a range from 600 to 3500 cm^{-1}, depending on the identity of the two molecules competing for the proton.

Figure 1 presents a representative survey of the [A·H$^+$·B]·Ar vibrational spectra in a series of complexes that explores differences in the O-H$^+$-O motif, as well as situations in which the proton is trapped between dissimilar atoms (such as O-H$^+$-N and O-H$^+$-RG). In most cases, the bands associated with $\nu_{sp}(\parallel)$ (red) could be identified by a process of elimination based on the unambiguous identification of transitions arising from the flanking structures. The plausibility of these assignments was then checked against ab initio calculations of the harmonic spectra, which typically yielded one $\nu_{sp}(\parallel)$ band within several hundred wavenumbers of the identified structures. The failure of such calculations to recover these bands more accurately provides clear evidence that these motions are indeed quite anharmonic. Isotopic substitution was also used in two cases, (Me$_2$O)$_2$H$^+$/(Me$_2$O)$_2$D$^+$ and Me$_2$O·H$^+$·H$_2$O/Me$_2$O·D$^+$·D$_2$O, which confirmed the assignments derived from the scheme outlined above. Because of the special properties associated with the symmetrical system, we will return to the isotope-dependence of the (Me$_2$O)$_2$H$^+$ spectrum after discussion of the overall trends.

One qualitative aspect common to all of the spectra obtained for the cold, isolated [A·H$^+$·B] systems investigated here is that these one-photon Ar predissociation spectra are much simpler than those previously reported for these systems with the use of IRMPD (12–14) or obtained from condensed phase samples. A specific example is shown in Fig. 1F. The previously reported condensed phase spectrum (gray) and its convoluted Gaussian fits (blue) (11) are included for comparison. The much simpler Ar predissociation spectrum not only allows the identification of the features that are due to the vibrational zero-point level, but also establishes that the intrinsic $\nu_{sp}(\parallel)$ oscillator strength is not strongly diluted by

extensive mixing with the background levels or by dynamical couplings (13).

Relationship between ν_{sp} and proton affinity. The origin of such large variations in ν_{sp} displayed by the analogous neutral complexes (e.g., Me$_2$O·HCl) has been subjected to intensive analysis over the years. Pimentel and co-workers (20), for example, correlated the vibrational frequency trends with differences in the proton affinities (ΔPA) between the two bases, and Barnes (21) later extended this study to a myriad of neutral combinations. Zundel (22) and others (23) considered the infrared spectra of a number of systems in salt crystals, and found that although ν_{sp} was correlated with the magnitude of ΔPA (or the analogous ΔpK_a, where K_a is the acid dissociation constant), the ν_{sp} bands were also shown to depend on the distance between the atoms connected by the shared proton, which is in turn strongly affected by the overall crystal

structure. This dependence on crystal structure obscured the correlation between ν_{sp} and ΔPA. The critical role of ΔPA in interpreting band positions was not applied to proton-bound cationic complexes until 1996, when Bieske and co-workers observed a linear scaling over a small range (300 cm^{-1}) in several weakly bound X·H$^+$·RG clusters (24). Very recently, gas-phase spectra of two asymmetric systems [Me$_2$O·H$^+$·THF, where THF is tetrahydrofuran, and acetone·H$^+$·THF] (14) in the low-ΔPA range (<25 kJ/mol) have been obtained using room-temperature IRMPD, in which assignments of $\nu_{sp}(\parallel)$ transitions among the overlapping bands (based on ab initio predictions) suggested a linear dependence on ΔPA. By eliminating the complicating influences of external forces and internal excitation at elevated temperatures, our technique offers a clear picture of the intrinsic molecular physics at play in the intermolecular proton bond.
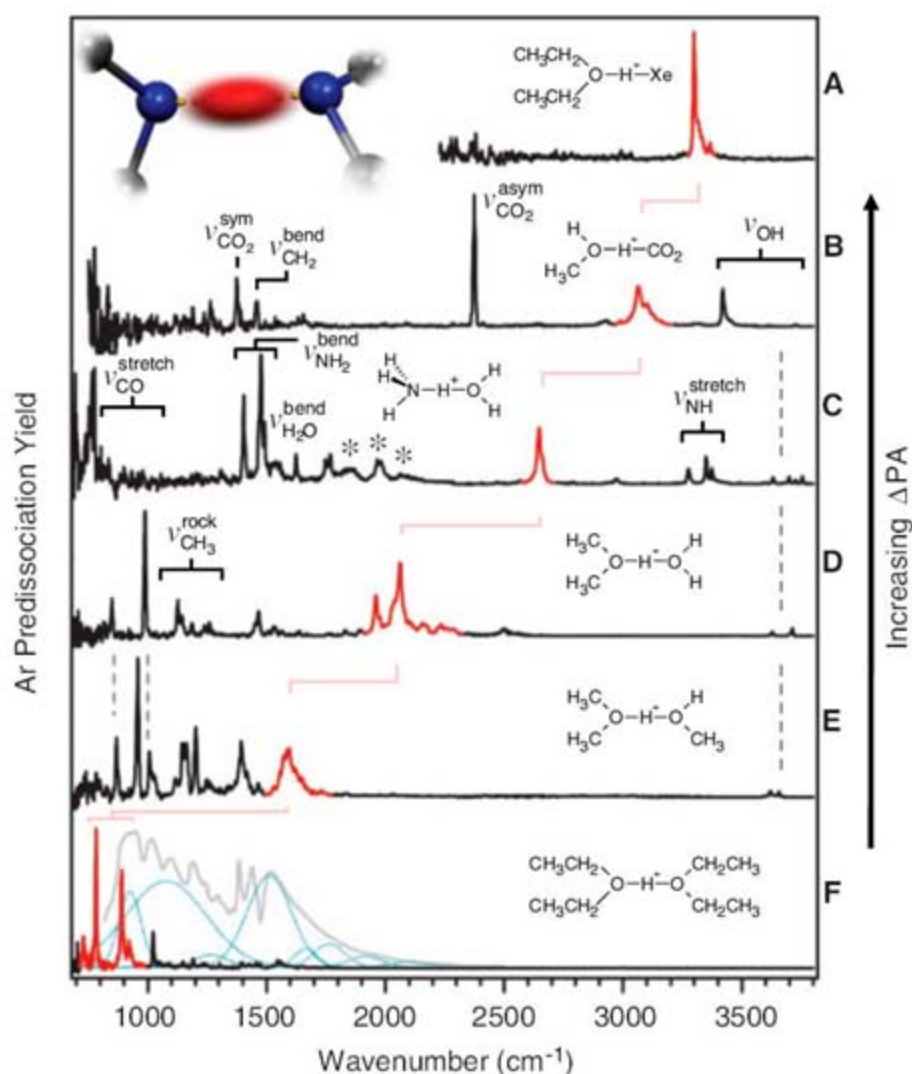

Fig. 1. Vibrational predissociation spectra of the [A·H$^+$·B]·Ar complexes for the following combinations: (**A**) Et$_2$O·H$^+$·Xe, (**B**) MeOH·H$^+$·CO$_2$, (**C**) NH$_3$·H$^+$·H$_2$O, (**D**) Me$_2$O·H$^+$·H$_2$O, (**E**) Me$_2$O·H$^+$·MeOH, and (**F**) Et$_2$O·H$^+$·Et$_2$O. The traces are presented in order of decreasing ΔPA (25) of the [A·H$^+$·B] complexes. Regions associated with vibrations more localized on the flanking molecules are indicated by ν_i^j labels, where i is the molecular moiety and j describes the characteristic motions involved in the vibrational mode. Bands indicated by the asterisk in (C) are due to an (H$_2$O)$_2^+$ impurity. The gray spectrum overlaying the trace in (F) corresponds to the solution phase spectrum, and the blue dashed lines indicate the Gaussian fits (11) used to extract the ν_{sp}-based transition energies reported for this complex. [Reprinted in part with permission from (11). Copyright 2006, American Chemical Society.] The vertical dashed gray lines in (C), (D), and (E) track features that persist over several complexes.

Figure 2 presents the ΔPA dependence of the centroid frequencies of the $\nu_{sp}(\|)$-based bands extracted from the spectra in Fig. 1, as well as those from 12 other systems that we studied. The relevant band positions are collected in Table 1 along with the ΔPA values for each pair of constituent molecules (25). The correlation between the absolute $\nu_{sp}(\|)$ position and ΔPA is thus revealed over a range of 2700 cm^{-1}, with all of the systems conforming closely to a universal trend, even though different atomic partners are involved in this data set.

It is useful to consider the qualitative origin of the large observed spectral shifts in the context of a simplified, one-dimensional potential for shared-proton motion, during which the heavy atoms are fixed and a proton is scanned between them. The resulting calculated curves for three representative systems are included in Fig. 2. At large values of ΔPA (Fig. 2C), the relative energies of the two possible proton transfer situations, AH$^+$·B and A·H$^+$B, are very different. This disparity localizes the proton to one side of the complex, effecting a high $\nu_{sp}(\|)$ frequency (>3000 cm^{-1}). As the relative energies become similar (Fig. 2, A and B), the proton transfer potential widens markedly and shifts $\nu_{sp}(\|)$ to lower energies (19).

In symmetrical systems with strong bases like ammonia, the trapping potential is calculated to develop a barrier at the midpoint, resulting in a double minimum shape (26). Earlier work on the NH$_3$·H$^+$·NH$_3$ system (26) in the free NH stretching region established that the two NH$_3$ groups are equivalent. This indicates that the vibrational zero-point lies above the barrier, thus

enabling the proton to be equally shared between the two N atoms. An analogous situation was also encountered in the highly basic anionic complex OH$^-$·H$^+$·OH$^-$, which displayed an isolated $\nu_{sp}(\|)$ transition at 670 cm^{-1} (5), lower in energy than any of the bands recovered in the cationic systems that we studied. Unfortunately, our initial survey of the NH$_3$·H$^+$·NH$_3$ spectrum in the low-energy range suggests that its $\nu_{sp}(\|)$ falls below 600 cm^{-1}, the lower limit of the apparatus.

Because the strongly basic systems develop a barrier in the shared-proton potential, we do not anticipate that there is an intrinsic asymptotic value for $\nu_{sp}(\|)$ at ΔPA = 0. Rather, in the limit that the first two vibrational levels fall below the barrier, the $\nu_{sp}(\|)$ transition energy evolves into a tunneling splitting. These considerations suggest that the correlation displayed in Fig. 2 reflects a more limited range of behavior where the shared-proton potential is flattened (Fig. 2B) but does not develop a substantial barrier. Characterization of the $\nu_{sp}(\|)$ band structure in high barrier systems thus presents a challenge for future work in this area.

Impact of complexation on flanking group vibrations. In the process of identifying discrete transitions associated with $\nu_{sp}(\|)$, we have also obtained sharp band patterns derived from the skeletal motions of flanking organic groups. These data provide an opportunity to understand how the pendant molecules are affected by the intermolecular proton bond. Our calculations suggest that, although the normal modes of vibration are collective motions of the entire complex, it is often the case that they are largely intramolecular

motions associated with one of the components. We highlight the case of the Me$_2$O·H$^+$·MeOH complex in Fig. 3. Figure 3G presents the observed spectrum in the fingerprint region; the left and right panels below it pertain to the transitions expected for the constituent molecules. Intuitively, one anticipates that the spectral signature associated with the skeletal motions of each monomer subunit should fall between the limiting patterns corresponding to the neutral (i.e., Me$_2$O or MeOH) and protonated (i.e., Me$_2$OH$^+$ or MeOH$_2^+$) species. Although the spectra of the neutral molecules are well known (27), this information was not available for the protonated analogs; we therefore obtained the spectra of the protonated analogs using Ar predissociation.

The traces in Fig. 3, A and B, present the IR spectra of the independent neutral components [Me$_2$O (Fig. 3A) and MeOH (Fig. 3B)], with the corresponding spectra of their protonated analogs displayed in the middle traces [Me$_2$OH$^+$ (Fig. 3C) and MeOH$_2^+$ (Fig. 3D)]. These spectra can be confidently assigned with the aid of ab initio calculations at the scaled harmonic level, because they are covalently bound systems that do not display large amplitude motion or unusual anharmonicity. In both neutral molecules (Fig. 3, A and B), the bands lowest in energy arise from the C-O stretching vibrations, and they are each shifted to lower energy (red-shifted) by about 200 cm^{-1} upon protonation. The calculations predict that two of the low-energy bands in the Me$_2$O·H$^+$·MeOH complex are due to the C-O stretches on the Me$_2$O group, which undergo about half the red-shift displayed by the analogous modes in the protonated ether. Similarly, the band between these two can be assigned to the C-O stretch primarily localized on
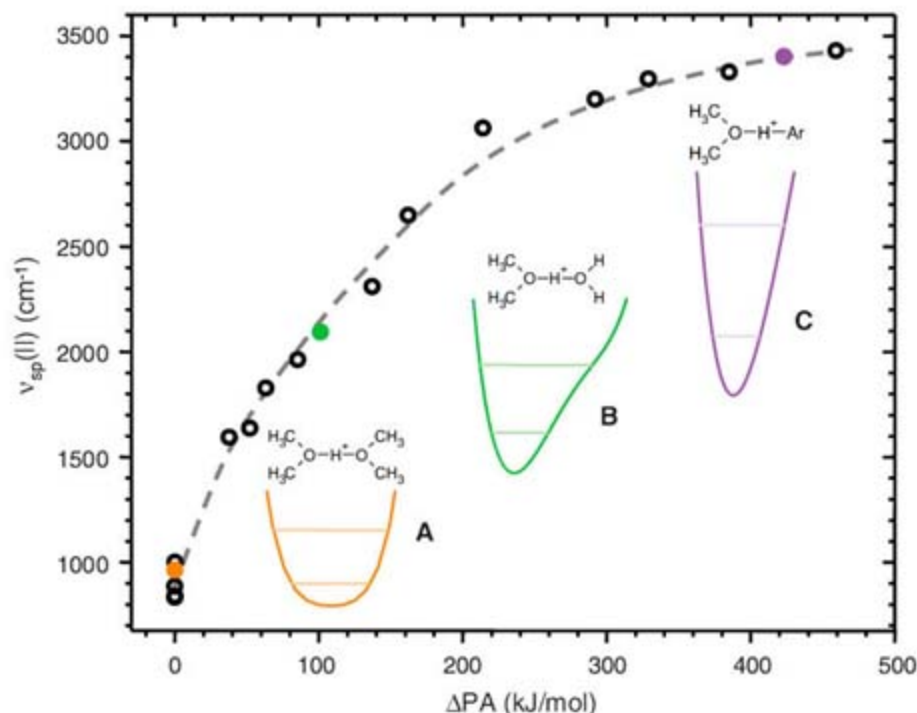

Fig. 2. Dependence of the $\nu_{sp}(\|)$-derived band locations on the difference in proton affinities (ΔPA) for all systems reported in Table 1 (25). (**A** to **C**) Calculated potential curves (MP2/aug-cc-pVDZ) generated by scanning the shared proton between the heavy atoms while fixing the exomolecular structures at the equilibrium geometry of the complex (28). Energy levels were derived by solving the one-dimensional Schrödinger equation for the first two vibrational eigenstates. The potential curves are color-coded to match colored points on the graph.

Table 1. Observed centroids (cm^{-1}) of shared-proton asymmetric stretching bands, $\nu_{sp}(\|)$, and relative (gas-phase) proton affinity (ΔPA) values for [A·H$^+$·B] complexes (25).

Complex	$\nu_{sp}(\|)$ (cm^{-1})	ΔPA (kJ/mol)
Et$_2$O·H$^+$·OEt$_2$	843	0
Me$_2$O·H$^+$·OMe$_2$	952	0
EtOH·H$^+$·HOEt	840	0
MeOH·H$^+$·HOMe	887	0
H$_2$O·H$^+$·OH$_2$	1002	0
Me$_2$O·H$^+$·HOMe	1595	38
Et$_2$O·H$^+$·HOEt	1638	52
MeOH·H$^+$·OH$_2$	1828	63
EtOH·H$^+$·OH$_2$	1964	85
Me$_2$O·H$^+$·OH$_2$	2094	101
Et$_2$O·H$^+$·OH$_2$	2310	137
H$_2$O·H$^+$·NH$_3$	2649	162
CO$_2$·H$^+$·HOMe	3064	214
Me$_2$O·H$^+$·Xe	3200	292
Et$_2$O·H$^+$·Xe	3296	329
MeOH·H$^+$·Ar	3330	385
Me$_2$O·H$^+$·Ar	3403	423
Et$_2$O·H$^+$·Ar	3431	459

MeOH, which likewise absorbs at a frequency intermediate between the bands in the neutral and protonated species.

In the Me$_2$O system, the CH$_3$ rocking band at 1200 cm^{-1} is completely suppressed upon protonation (compare Fig. 3, A and C), but it is recovered in the spectrum of the bridging complex. This behavior is reminiscent of the band disappearance alluded to earlier in the symmetrical complexes (10, 11). In the case of the CH$_3$ rocks, however, their suppression is clearly a consequence of the inherent reduction in the transition moments associated with the normal mode fundamentals rather than heterogeneous broadening due to spectral diffusion or some other dynamical process.

A persistent feature at 1450 cm^{-1} associated with the CH$_2$ bending motion appears in all of the spectra and seems to be largely decoupled from the bonding environment at the oxygen center. These CH$_2$ bends are suppressed upon complexation, but nevertheless remain sharp and distinct features. The shared-proton transition in Me$_2$O·H$^+$·MeOH lies higher in energy at ~1600 cm^{-1}, about 500 cm^{-1} below the predicted value at the harmonic level. This large anharmonicity undoubtedly results from the distorted "shelf-like" potential governing the shared proton in this asymmetric complex (Fig. 2B). Consequently, the $\nu_{sp}(\parallel)$ transition can be much more accurately recovered by numerically integrating the

Schrödinger equation for a one-dimensional potential slice along the proton transfer coordinate. This band is the broadest in the spectrum, appearing just above the bending fundamental associated with the dangling OH on the MeOH moiety.

Analysis of symmetrical complexes. We turned next to the homogeneous (symmetrical) complexes, for which there is a considerable data set available from previous studies carried out on solution-phase as well as crystalline samples (23). On the basis of extensive work recently focused on the prototypical H$_5$O$_2^+$ system, the shared-proton motion is often associated with a strong doublet near 1000 cm^{-1} and a second band in the 1600 to 1800 cm^{-1} range, whose intensity is much more dependent on composition of the solution in the condensed phase (11), as well as on the identity of the rare-gas atom in the gas-phase studies (6). The simple doublet nature of the 1000-cm^{-1} band has proven very difficult to recover in high-level theoretical treatments of this system (e.g., many-dimensional vibrational coupling on a full potential surface) (7). In spite of the subtlety associated with this fine structure, a doublet character has been extracted from fits of the asymmetric, broad bands displayed by the symmetrical complexes in the condensed phase (11). The assignment of the band at higher energy (~1800 cm^{-1}) is uncertain, and may well be different in the condensed phase given that the structures of the complexes are thought to

change from a staggered orientation of the flanking groups in isolation to pseudo-planar in the presence of an anion support (11). This structural change also raises the specter that the lower energy envelopes recovered in solution similarly may not reflect the intrinsic nature of the intermolecular proton bond. These uncertainties warrant a closer look at one of the isolated symmetrical systems.

Figure 4 shows the results of the isotopic substitution study of the Me$_2$O·H$^+$·Me$_2$O system, in which we independently recorded the spectra of the Me$_2$O·H$^+$·Me$_2$O (Fig. 4B) and Me$_2$O·D$^+$·Me$_2$O (Fig. 4C) complexes. Most important, the three strong bands in the H isotopologue centered around 900 cm^{-1} are quenched upon H/D substitution, and a closely spaced doublet emerges at ~700 cm^{-1}. Arrows in Fig. 4B indicate the locations of the C-O stretching levels in Me$_2$OH$^+$, which suggest that, as in the asymmetric complexes (Fig. 3), the bands near 900 cm^{-1} involve this motion. What is surprising, therefore, is that these bands disappear when the shared H is replaced by D, indicating that the shared-proton motion is somehow primarily responsible for all three of the strong transitions in the H species. This behavior is particularly notable in the context of the observations in the condensed phase that bands associated with the organic groups disappear upon complexation in the proton-bound dimers (10, 11). Moreover, similar

Fig. 3. Decomposition of the Me$_2$O·H$^+$·MeOH infrared spectrum into the dominant contributions from each of the constituent molecules with the aid of the independent spectra of (**A**) Me$_2$O, (**B**) MeOH, (**C**) Me$_2$OH$^+$, and (**D**) MeOH$_2^+$. Neutral spectra are displayed as Gaussians about the vibrational origins with heights corresponding to the observed relative intensities of the reported room-temperature spectra (27). (**E** and **F**) The ether-based (teal) and alcohol-based (blue) contributions to the overall spectrum of the complex (**G**). The locations of the backbone transitions in (E) and (F) are intermediate between their analogous positions in the protonated [(C) and (D)] and neutral [(A) and (B)] species, showing the delocalized character of the shared proton, even in the asymmetric complex.
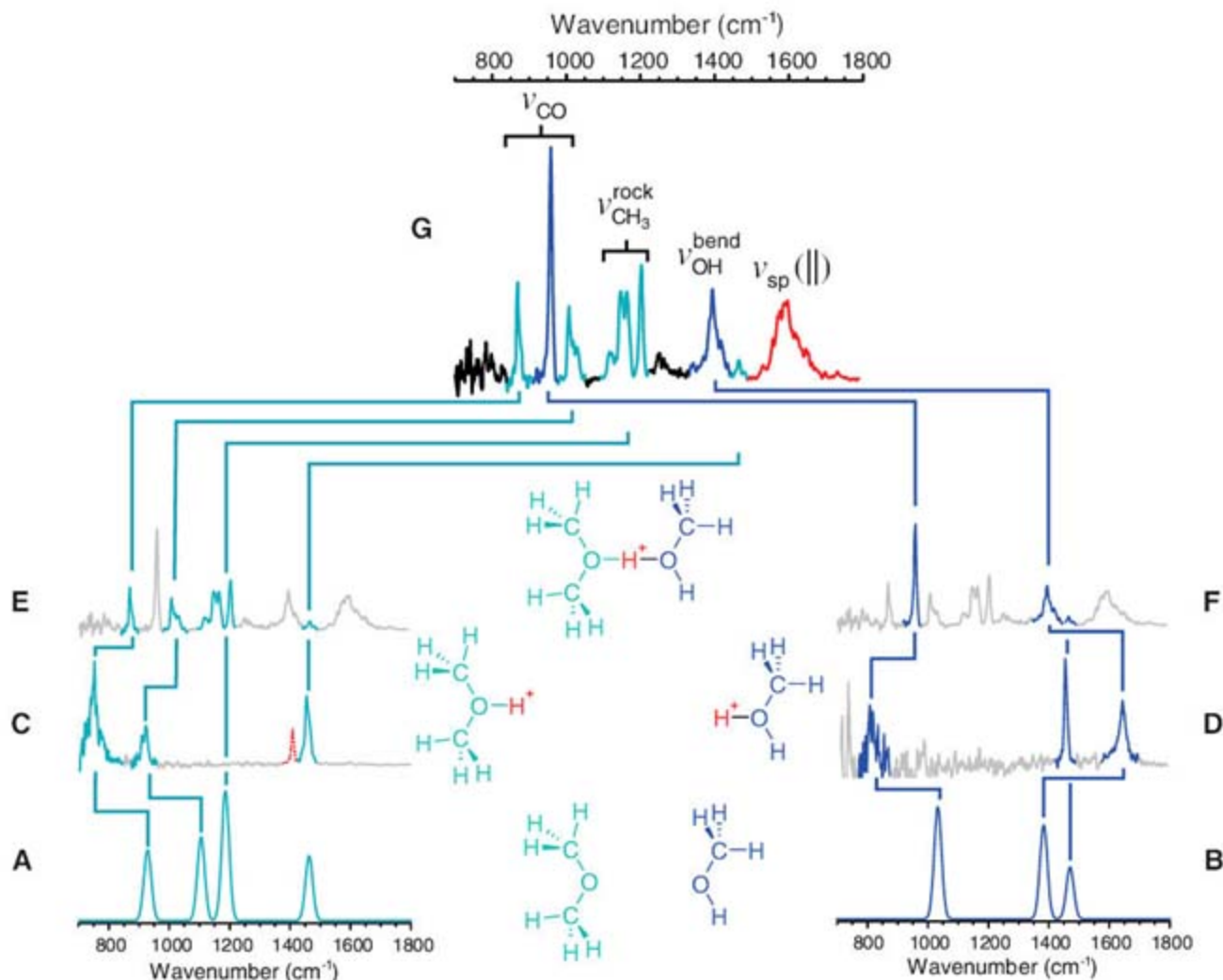
behavior is well established in the neutral acid dimers [such as (RCO$_2$H)$_2$].

To explain the curious suppression of the intramolecular transitions in the symmetrical systems, Stoyanov has invoked a spectral diffusion mechanism in the condensed media, whereby the large amplitude motion of the shared proton effectively modulates the transitions of flanking groups such that they broaden and disappear in the background (10, 11). The clear example of similar band suppression in the isolated Me$_2$O·H$^+$·Me$_2$O system upon deuteration, however, allows us to gain a deeper understanding of its origin. In particular, because the isolated system is accessible with ab initio calculations that can be directly compared with experiment, we can explore whether the curious intensity behavior is a low-order phenomenon or whether it requires high-level, full-dimensional quantum treatment (such as the situation encountered in H$_5$O$_2^+$) to reveal the qualitative motions underlying the observed doublet pattern.

Moore et al. (13) previously reported ab initio (cc-pVDZ basis at MP2 level) harmonic spectra for the (Me$_2$O)$_2$H$^+$ complex in the context of the IRMPD spectra. To determine the expected isotope-dependence of the IR fundamentals and extract the normal mode displacement vectors, we carried out calculations at the B3LYP/6-311++G (d,p) level (28), which yielded a similar pattern of bands with comparable intensities. Fig. 4, A and D, presents the harmonic spectra for the two isotopologues, which recover the character of the overall band patterns surprisingly well given the crude level of treatment. The displacement vectors

(insets) associated with the strongest two bands in the H isotopologue involve substantial C-O stretching motions. As such, we can regard these vibrations as being derived from mixed combinations of the shared proton with the out-of-phase collective motion of the two symmetric C-O stretches. There are actually three such parallel transitions whose intensities originate, in this zero-order picture, from the two symmetric C-O stretches (one on each monomer) and v_{sp}(‖), with the third (the in-phase symmetric C-O stretch) accounting for very little oscillator strength. This assignment is consistent with the earlier conclusions of Moore et al., in their ab initio study of this system (13).

Turning to the heavy isotope, we anticipated that this C-O/v_{sp}(‖) mixing scheme would decouple as the v_{sp}(‖) value dropped out of range in the D isotopologue, implying that the usual two C-O stretching bands should emerge at higher energy relative to the v_{sp}(‖)-based transition. Such levels are calculated to occur in the 900-cm^{-1} region, but the oscillator strengths for their IR fundamentals are predicted to be much smaller than the intrinsic strength of the v_{sp}(‖) transition (<1%), making them difficult to observe, given the noise in the experimental spectrum. These observations thus point to an alternative explanation for band suppression that does not require spectral diffusion, but rather emphasizes the extreme disparity in the absorption cross-sections for the intramolecular vibrations relative to that of v_{sp}(‖). In this context, the more curious case is the multiplet structure of v_{sp}(‖) in the light isotopologue. We suggest that, rather than being suppressed, this structure reflects

a marked intensity enhancement of the C-O stretching motions through their interaction with v_{sp}(‖) when the two modes are nearly degenerate. In this model, the multiplet results from interaction with the flanking groups, rather than reflecting the intrinsic signature of v_{sp}(‖). These cursory observations at the harmonic level clearly highlight the need for further theoretical investigation of the band suppression effect in order to treat the expected large anharmonicities at play in these relatively floppy systems.

Notably, not all of the bands on the flanking molecules are suppressed, as indicated by the weaker C-H stretching transitions at the high-energy end of the spectra (Fig. 4). These envelopes are very well recovered by the calculated harmonic spectra (Fig. 4, A and D), and they occur with intensities that are both consistent with theoretical expectations and similar to those of the bands in protonated dimethyl ether (Me$_2$OH$^+$).

Both of the proton-bound ethers studied here [(R$_2$O)$_2$H$^+$, where R = Me, Et] are dominated by sharp multiplets in the low-energy range, and our analysis supports the conclusion that this structure arises from mixing between v_{sp}(‖) and the C-O stretches (13). One way to further explore this behavior within the context of symmetrical systems is through substitution of one of the alkyl groups on each oxygen atom with an H atom, a modification readily accommodated by synthesizing the proton-bound alcohols, (MeOH)$_2$H$^+$ and (EtOH)$_2$H$^+$. Figure 5 compares the finger-

Fig. 4. Isotope substitution study of the symmetrical Me$_2$O·H$^+$·Me$_2$O complex: (**A**) calculated spectrum and (**B**) Ar-tagged vibrational predissociation spectrum of Me$_2$O·H$^+$·Me$_2$O; (**C**) predissociation spectrum and (**D**) calculated spectrum of Me$_2$O·D$^+$·Me$_2$O. Calculations are carried out at the B3LYP/6-311++G(d,p) level with the CH stretches scaled by 0.967, the factor required to match the observed values for the neutral molecules (28). Bands predominantly assigned to v_{sp}(‖) activity are indicated in red. Included in insets are the normal mode displacements for the highlighted transitions of interest, showing that although the shared proton is coupled to the CO stretches in both cases, the resulting band patterns of the isotopologues are notably different. The CO stretches of the protonated Me$_2$O species are indicated by black arrows. The gray trace in (B) is the previously reported IRMPD spectrum of bare Me$_2$O·H$^+$·Me$_2$O (13). The peaks labeled with asterisks in (C) are due to the overlapping ^{13}C isotopologue of the H-labeled species.
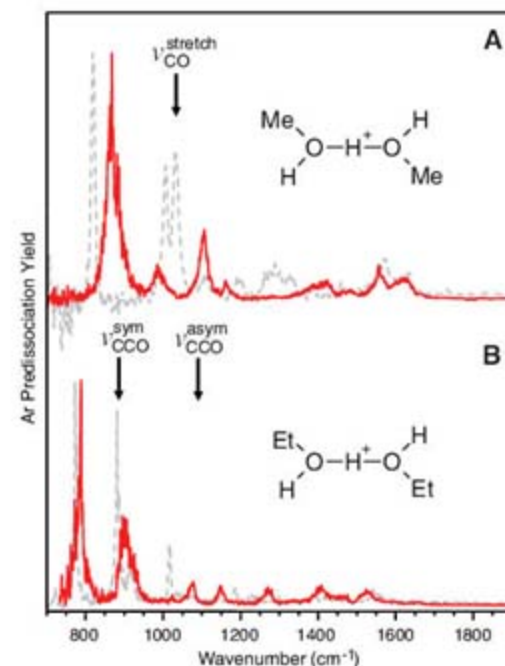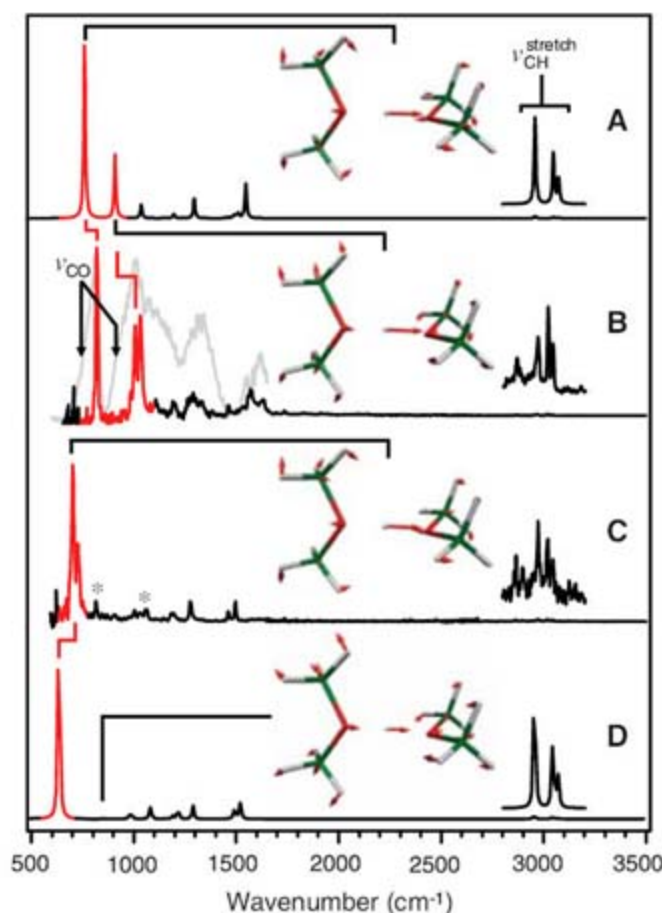

Fig. 5. Comparison of the vibrational predissociation spectra of the homogeneous dimers: (**A**) Me$_2$O·H$^+$·Me$_2$O·Ar (gray, dashed) and MeOH·H$^+$·MeOH·Ar (red, solid), and (**B**) Et$_2$O·H$^+$·Et$_2$O·Ar (gray, dashed) and EtOH·H$^+$·EtOH·Ar (red, solid). Arrows indicate the C-O stretching fundamentals corresponding to the neutral constituent molecules where $v_{CO}^{stretch}$ indicates the CO stretch in methanol (30), whereas v_{CCO}^{sym} and v_{CCO}^{asym} denote the symmetric and asymmetric CCO stretching motions of ethanol, respectively (31).

print region spectra of the alcohols [methanol (Fig. 5A) and ethanol (Fig. 5B)] with their respective ethers. The lowest energy band obtained here for $(MeOH)_2H^+$ is in close agreement with that recovered previously by Fridgen et al. in an IRMPD study of the bare complex, whereas that of the Ar-tagged $(EtOH)_2H^+$ spectrum is clearly sharper and higher in energy than that observed through IRMPD (29). At the qualitative level, the main difference found upon tagging is a marked reduction in the diffuse bands at higher energy.

Although the alcohols display strong features in the region of 800 to 1000 cm^{-1}, close to those found in their respective ether complexes, the bands in the alcohols are clearly broadened. The alcohols introduce complications that could explain such broadening; for example, there are two low-lying conformers (with different calculated band patterns) that depend on the proximity of the hydrogen atoms (29). Other factors include the different orientation of the C-O stretching displacements relative to the more parallel motion in the ethers, and the possibility of mixing with the nearby H^+-O-H bending motion of the dangling hydrogen atom.

The dominant features in the methanol complex (Fig. 5A) are blue-shifted relative to those in $(Me_2O)_2H^+$, and each strong band appears with a weaker absorption on the high-energy side. In contrast, the two most intense bands in the ethanol complex (Fig. 5B) fall quite close to those in $(Et_2O)_2H^+$. The lowest energy band (at 788 cm^{-1}) is the sharpest $v_{sp}(\parallel)$-derived feature obtained from either alcohol. The sharpening of this band in the larger system establishes that the widths do not simply reflect the density of background states, which would be largest in the R = Et case. The rather different behavior of the two alcohols emphasizes that the detailed distribution

of the oscillator strength associated with $v_{sp}(\parallel)$ in the symmetrical complexes is system-dependent.

Implications of the infrared spectra. Across the range of 18 complexes surveyed here, the dominant oscillator strengths traced to the shared-proton vibrational fundamentals evolve over a 2700-cm^{-1} span. The transitions are isolated and sharp when the vibrational quantum exceeds 2000 cm^{-1}, but they broaden somewhat and acquire multiplet structure as the shared-proton vibration merges with those arising from skeletal motions of the organic scaffolds in the fingerprint region. The shared proton–based band centers are observed to be closely correlated with the difference in proton affinities of the two partner molecules. The well-resolved spectral features in the fingerprint region also allow us to follow the deformations suffered by the monomer subunits as they become bound by the proton. This aspect is demonstrated in the detailed study of the $Me_2O \cdot H^+ \cdot MeOH$ species.

References and Notes

1. W. W. Cleland, M. M. Kreevoy, Science **264**, 1887 (1994).
2. H. Luecke, H.-T. Richter, J. K. Lanyi, Science **280**, 1934 (1998).
3. T. E. DeCoursey, Physiol. Rev. **83**, 475 (2003).
4. F. Garczarek, K. Gerwert, Nature **439**, 109 (2006).
5. E. G. Diken, J. M. Headrick, J. R. Roscioli, J. C. Bopp, M. A. Johnson, J. Phys. Chem. A **109**, 1487 (2005).
6. N. I. Hammer et al., J. Chem. Phys. **122**, 244301 (2005).
7. A. B. McCoy, X. Huang, S. Carter, M. Y. Landeweer, J. M. Bowman, J. Chem. Phys. **122**, 061101 (2005).
8. X. Huang, B. J. Braams, J. M. Bowman, J. Chem. Phys. **122**, 044308 (2005).
9. M. Kaledin, A. L. Kaledin, J. M. Bowman, J. Phys. Chem. A **110**, 2933 (2006).
10. E. S. Stoyanov, Phys. Chem. Chem. Phys. **2**, 1137 (2000).
11. E. S. Stoyanov, C. A. Reed, J. Phys. Chem. A **110**, 12992 (2006).
12. K. R. Asmis et al., Science **299**, 1375 (2003).
13. D. T. Moore et al., ChemPhysChem **5**, 740 (2004).
14. T. D. Fridgen et al., Phys. Chem. Chem. Phys. **7**, 2747 (2005).
15. G. Scoles, K. K. Lehmann, Science **287**, 2429 (2000).
16. M. A. Johnson, W. C. Lineberger, in Techniques for the Study of Ion-Molecule Reactions, J. J. M. Farrar, W. H. Saunders, Ed. (Wiley, New York, 1988), vol. XX, pp. 591–635.
17. M. Okumura, L. I. Yeh, J. D. Myers, Y. T. Lee, J. Chem. Phys. **85**, 2328 (1986).
18. J. M. Headrick, J. C. Bopp, M. A. Johnson, J. Chem. Phys. **121**, 11523 (2004).
19. J. R. Roscioli, E. G. Diken, M. A. Johnson, J. Phys. Chem. A **110**, 4943 (2006).
20. B. S. Ault, E. Steinback, G. C. Pimentel, J. Phys. Chem. **79**, 615 (1975).
21. A. J. Barnes, J. Mol. Struct. **100**, 259 (1983).
22. G. Zundel, Adv. Chem. Phys. **111**, 1 (2000).
23. A. Novak, in Structure and Bonding (Springer-Verlag, Berlin, 1974), pp. 177–216.
24. S. A. Nizkorodov, O. Dopfer, M. Meuwly, J. P. Maier, E. J. Bieske, J. Chem. Phys. **105**, 1770 (1996).
25. E. P. Hunter, S. G. Lias, "Proton affinity evaluation," in NIST Chemistry WebBook, NIST Standard Reference Database Number 69, P. J. Linstrom, W. G. Mallard, Eds. (National Institute of Standards and Technology, Gaithersburg, MD, 2005); available online (http://webbook.nist.gov).
26. J. M. Price, M. W. Crofton, Y. T. Lee, J. Phys. Chem. **95**, 2182 (1991).
27. T. Shimanouchi, "Molecular vibrational frequencies," in NIST Chemistry WebBook, NIST Standard Reference Database Number 69, P. J. Linstrom, W. G. Mallard, Eds. (National Institute of Standards and Technology, Gaithersburg, MD, 2005); available online (http://webbook.nist.gov).
28. M. J. Frisch et al., Gaussian 03 (Gaussian Inc., Wallingford, CT, 2003).
29. T. D. Fridgen, L. MacAleese, T. B. McMahon, J. Lemaire, P. Maitre, Phys. Chem. Chem. Phys. **8**, 955 (2006).
30. A. J. Barnes, H. E. Hallam, T. Faraday Soc. **66**, 1920 (1970).
31. A. J. Barnes, H. E. Hallam, T. Faraday Soc. **66**, 1932 (1970).
32. We thank the Experimental Physical Chemistry Division of the NSF and the donors of the American Chemical Society Petroleum Research Fund for partial support of this work. We also thank J. Headrick for advances in the laser capabilities in our laboratory.

Structure of Fungal Fatty Acid Synthase and Implications for Iterative Substrate Shuttling

Simon Jenni,* Marc Leibundgut,* Daniel Boehringer, Christian Frick, Bohdan Mikolásek, Nenad Ban†

We report crystal structures of the 2.6-megadalton $\alpha_6\beta_6$ heterododecameric fatty acid synthase from Thermomyces lanuginosus at 3.1 angstrom resolution. The α and β polypeptide chains form the six catalytic domains required for fatty acid synthesis and numerous expansion segments responsible for extensive intersubunit connections. Detailed views of all active sites provide insights into substrate specificities and catalytic mechanisms and reveal their unique characteristics, which are due to the integration into the multienzyme. The mode of acyl carrier protein attachment in the reaction chamber, together with the spatial distribution of active sites, suggests that iterative substrate shuttling is achieved by a relatively restricted circular motion of the carrier domain in the multifunctional enzyme.

Fatty acid synthesis is a central metabolic pathway that entails the iterative elongation of fatty acid chains through a set of chemical reactions conserved in all kingdoms of life. Whereas in most bacteria and plants, synthesis is carried out by sets of isolated enzymes (type II) (1, 2), fungi and vertebrates have evolved large, although architecturally diverse, multifunctional synthases (type I) (3–6). Mammalian fatty acid synthase (FAS) is a 540-kD α_2 homodimer (7, 8), whereas fungal FAS is a 2.6-megadalton $\alpha_6\beta_6$ heterododecameric complex (9–12). Despite their fundamentally different architectures, both types of megasynthases represent independent biosynthetic factories, because they integrate all necessary enzymatic activities together with acyl carrier protein (ACP) domains used for covalent substrate shuttling from one active site to the other.

In fungi, acetyl primer and malonyl elongation substrates are loaded from coenzyme A (CoA) to ACP by acetyltransferase (AT) and malonyl/palmitoyl transferase (MPT) and condensed to acetoacetyl-ACP in a decarboxylative reaction catalyzed by ketoacyl synthase (KS). In three subsequent reaction steps, the β-carbon groups are processed by ketoacyl reductase (KR), dehydratase (DH), and enoyl reductase (ER), which results in fully saturated acyl-ACP that can serve directly as a primer for the next

Institute of Molecular Biology and Biophysics, ETH Zurich, 8092 Zurich, Switzerland.

*These authors contributed equally to this work.
†To whom correspondence should be addressed E-mail: ban@mol.biol.ethz.ch

condensation reaction. In each reaction cycle, the growing acyl-chain is elongated by two carbon units until C_{16} or C_{18} fatty acids are back-transferred to CoA by MPT (6).

Here, we describe high-resolution structures of the fungal type I FAS from *Thermomyces lanuginosus* and of a complex with the nicotinamide adenine dinucleotide phosphate (NADP+) cofactor. We also present a structure-based model for substrate shuttling in fungal FAS, which may serve as a paradigm for mechanistically related multienzyme complexes.

Structure determination. We obtained experimentally phased electron density maps of *T. lanuginosus* FAS at 3.1 Å resolution by domainwise electron density averaging of the six copies in the asymmetric unit (13) (fig. S1). The refined atomic models include 21,127 amino acids (89% of the total number of residues), 6 flavin mononucleotides (FMN), and 12 NADP+ molecules, in the case of the NADP+-soaked crystal. In contrast to the structure of yeast FAS described in the accompanying paper (14), electron density corresponding to ACP was too weak

to allow us to build a model of this domain. The phosphopantetheine (PPT) transferase (PT) required for initial apo-ACP activation, for which we had considered the apex of the particle as a possible location (12), was not visible, and we presume that it is flexibly attached at the C terminus of the α chain.

Overall architecture of the fungal FAS. The overall structure of the fungal FAS can be roughly divided into three distinct units: the central wheel, which forms the equatorial disk of the particle, and two domes, which arch over the reaction
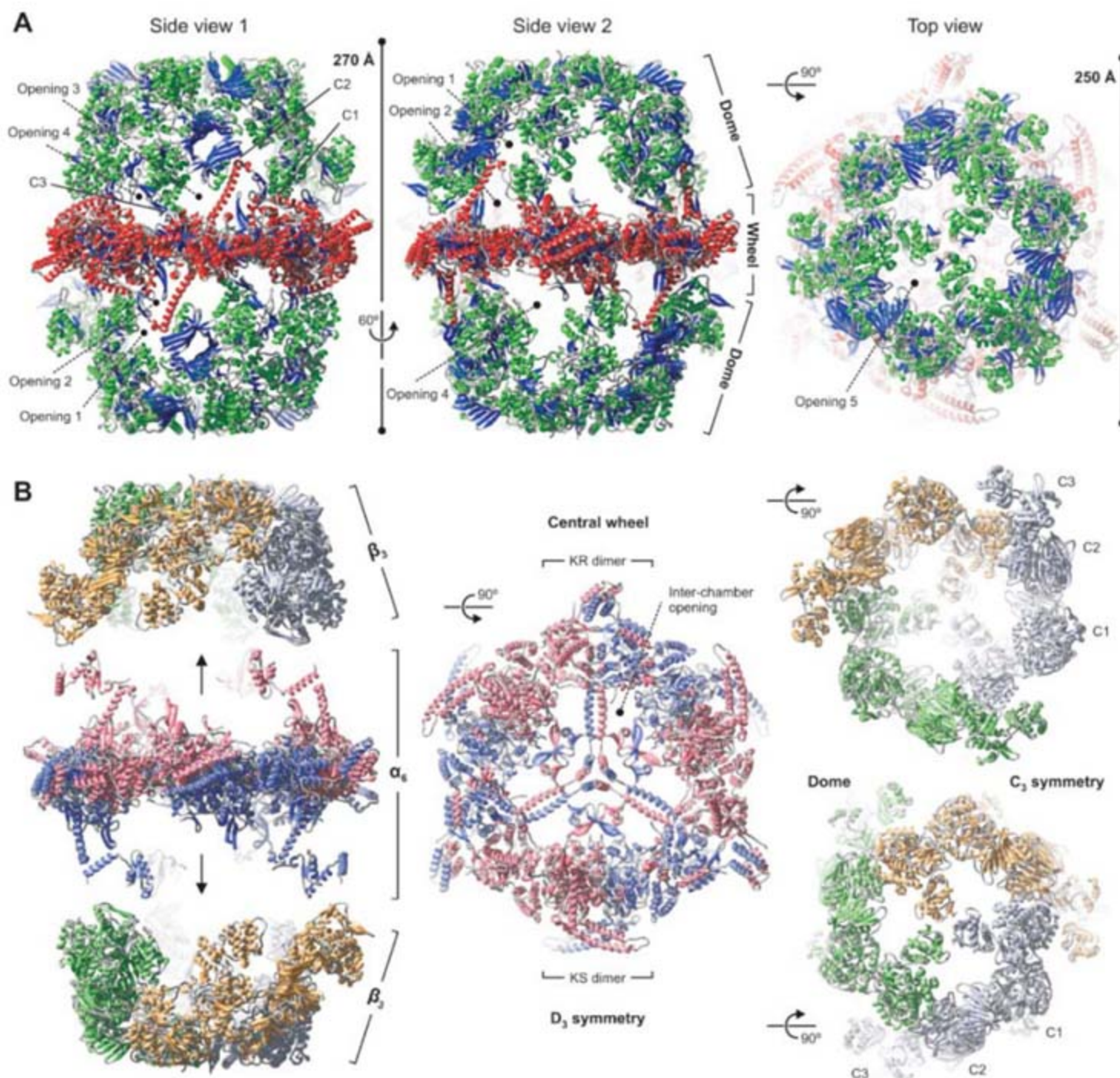
Fig. 1. Structure of the fungal FAS. (**A**) A ribbon representation of the refined *T. lanuginosus* FAS structure is shown in two side views along the two-fold axis of symmetry (left and middle) and in a top view along the three-fold axis of symmetry (right). α Helices belonging to the central wheel are colored red, α helices of the capping domes are shown in green, β strands are colored blue. The particle has an approximate height and width of 270 Å by 250 Å. "Windows" in the side walls and at the top of the particle are labeled as openings 1 to 5, connections between the dome and the central disk as C1, C2, and C3. (**B**) Quaternary structure and subunit distribution of the FAS complex. The heterododecameric complex is composed of a D_3-symmetrical α_6 hexamer (subunits are colored pink and blue) and two C_3-symmetrical β_3 trimers (subunits are colored green, yellow and gray). In the side view (left), the β_3 trimers have been moved away from the α chains as indicated by the arrows.

chambers (Fig. 1A). Five major, structurally unique openings or "windows" in the side walls and at the top provide access to the reaction chambers from the outside (Fig. 1A), and six openings in the central disk connect the upper and the lower reaction chambers (Fig. 1B).

The high-resolution structures confirm and substantially particularize the previously reported α- and β-subunit distribution within the FAS complex quaternary structure (9, 12). The complete backbone traces of the six α chains are highly intertwined and form the entire central wheel, which has D₃ symmetry. The apical dome has C₃ symmetry and is composed of three complete β chains and three times 94 N-terminal residues contributed from the α chains (Fig. 1B). The arches of the domelike superstructure rest on the α-chain platform by means of three distinct types of abutments or intersubunit connections (Fig. 1, A and B). The β chains are not

as extensively intertwined as the α chains, and individual subunits can be discerned within the assembly.

Catalytic core domains caged within expansion segments. The catalytic core domains, defined by the homology to the FAS type II enzyme structures, are embedded within expansion segments. These interspersed sequences are integral parts of the α and β chains in the form of intradomain insertions in all conserved catalytic folds, as well as additional, separate domains (Fig. 2A). They correspond to ~50% of the total mass of the complex and mediate the majority of architectural interactions by wrapping around, squeezing between, clamping and interconnecting the core domains (fig. S4). With the exception of the KS and KR dimer interfaces, there is minimal direct interaction between regions corresponding to bacterial core domains.

Although in some bacterial multifunctional enzymes, such as the fatty acid β-oxidation

complex (15), individual domains have been observed to form more tangential interactions, the expansion segments in fungal FAS stably integrate the catalytic domains into the complex. Through this assembly mode, the hollow structure of the multienzyme is generated from a small number of subunits, each displaying a high level of complexity. Compared with other macromolecular assemblies with large interior volumes, such as multimeric chaperones, proteases, and icosahedral viral capsids, fungal FAS is built out of fewer polypeptide chains and has more porous and permeable reaction-chamber walls. Analogous to FAS, openings through which substrates could enter into reaction chambers are also observed in pyruvate dehydrogenase (PDH) complexes (16).

Structure of the subunits. One 210-kD α and one 230-kD β chain combine to build one nonredundant portion of the fungal FAS complex (Fig. 2B). The complicated folds of the two unusually long polypeptides reveal numerous
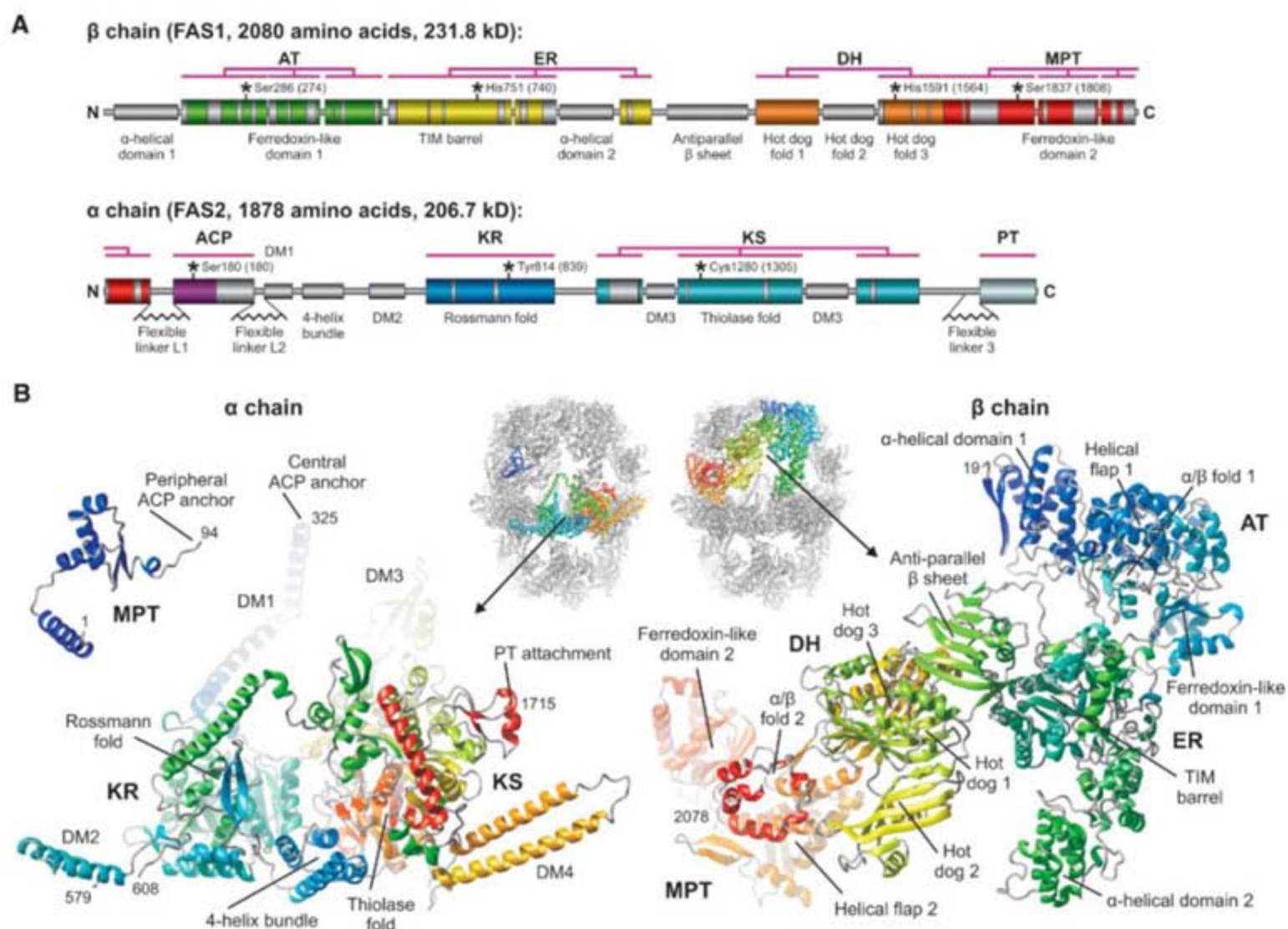
Fig. 2. Domain organization and structure of the subunits of fungal FAS. (**A**) Domain organization of the FAS α and β polypeptides from *T. lanuginosus*. Conserved sequence stretches structurally related to known enzymes are shown in color (AT, green; ER, yellow; DH, orange; MPT, red; ACP, purple; KR, blue; KS, cyan; PT, light blue). Functional domains are indicated by magenta bars. Structural domains, expansion segments, and interdomain linkers are colored gray. Asterisks indicate the position of catalytic amino acid residues (corresponding residues in yeast FAS are shown in parentheses, table S7) and zigzag lines the position of flexible linkers. (**B**) Structure of the α and β subunits. Ribbon representations of the α chain (left) and the β chain (right) are colored in rainbow from dark blue (N terminus) to red (C terminus).

intrachain contacts between residues distantly located in sequence (Fig. 3, E, F, and G). As described in more detail (supporting online text), the catalytic domains on the β chain are arranged in the same order as is observed in the primary structure. The appearance of the α chain is dominated by the KR and KS catalytic folds, from which extensions depart that provide interaction surfaces for oligomerization (Fig. 2B).

Both chains contain several solvent-exposed loops that may be accessible to proteases (fig. S5). In fact, because of proteolytic degradation during enzyme isolation procedures (*17*), early investigations of the yeast FAS from the late 1960s considered the multienzyme to be an associated complex of many individual proteins, analogous to bacterial systems (*6*, *18*).

Subunit interactions. The α chains in the central disk are intertwined and form numerous α-β subunit contacts (Fig. 1). Because of the D₃ symmetry of the α hexamer, α-α subunit contacts mainly involve homodimeric interfaces and interlocking modules arranged along the two-fold axis of symmetry. A large surface area of ~77000 Å² is buried when six α subunits assemble, to which the two highly conserved dimerization interfaces of the KR and KS (*8*, *15*, *19*, *20*) make substantial contributions.

Additional α-α contacts are mediated by four dimerization modules (DMs). Two extended α helices in the wheel associate in a head-to-head manner and constitute the thin spokes between the central hub and the KR domains (DM1) (Fig. 3A). The second type of spoke assembly (DM3) is formed by one of the sequence insertions in the KS core domain, where the expansion segments clamp together via a long double-stranded antiparallel β sheet capped with an α-β fold (Fig. 3C). Two peripheral DMs provide an elegant way to further stabilize the α-subunit association by essentially tying it together: The first (DM2) is formed by a four-helix bundle, which exhibits flexibility with respect to the body of the FAS (Fig. 3B); the second (DM4) is defined by

expansion segment helices at the outside of the KS (Fig. 3D).

The β-β contacts in the dome are restricted to helical domain 1, responsible for the formation of the trimeric apical connection of the arches (Fig. 3K), and to the interaction of the DH with the AT domain of a neighboring β chain. The construction of the dome buries less than ~5000 Å² surface area between the three β subunits, which is small compared with the observed contacts among the α chains, and allows slight motions of the individual domains.

The three different α-β connections C1, C2, and C3, where the arches depart from the equatorial disk (Fig. 1A), are all mediated by expansion segments. In C1, α-helical domain 2 beneath the ER core packs on one KS expansion segment, clamps the extended part of the KR-KS linker, and forms additional interactions with the long helix lying on the KS core structure (Fig. 3J). C2 is formed between the same α and β subunits as C1 and solely involves packing of an additional hot dog fold inserted in the
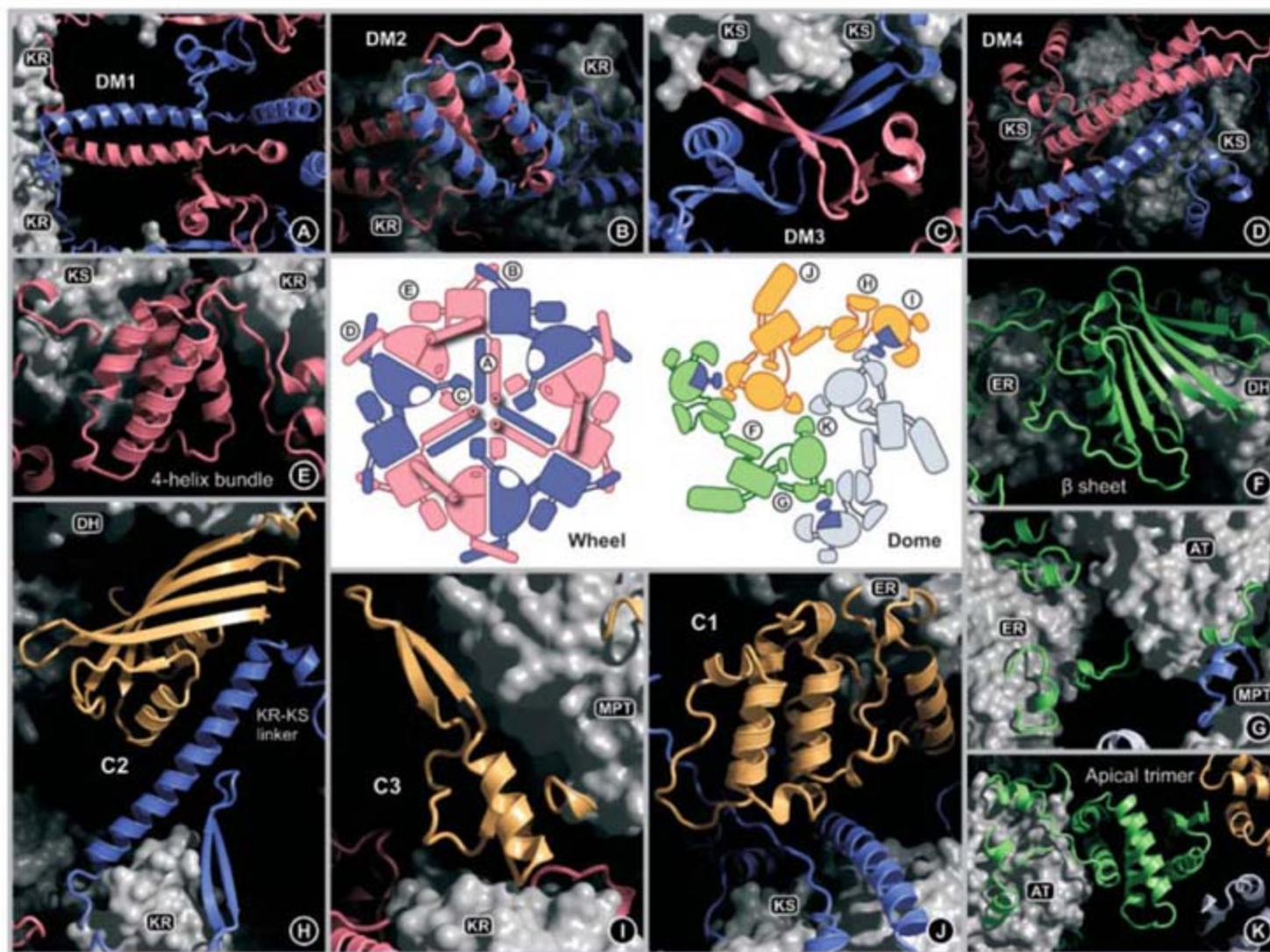
Fig. 3. Intra- and intersubunit interactions in the fungal FAS complex. Schematic topologies for the hexameric central wheel and the trimeric dome are shown in the middle. α Subunits are colored pink and blue, β subunits green, yellow, and gray. The labels (A) to (K) indicate the interaction motifs depicted in the surrounding panels. In the close-up views, catalytic core domains are shown as gray surfaces, and the expansion segments that mediate intra- and intersubunit interactions are shown as ribbons. (**A** to **D**) DMs of the central wheel. (**E**) Four-helix bundle within one α chain between KR and KS. (**F**) Antiparallel β sheet of the β chain between the ER and DH. (**G**) Contacts involving short expansion segments of AT, ER, and MPT. (**H** to **J**) Intersubunit α-β connections C1, C2, and C3. (**K**) Apical trimeric connection formed by the N-terminal α-helical domains of three β chains.

pseudodimeric DH against the long α helix of the KR-KS linker (Fig. 3H). Connection C3 is formed between an MPT expansion segment of the β subunit and a short insertion in the KR of a second α chain (Fig. 3I).

Double-anchored ACP in the reaction chamber. The electron density for the N-terminal region of the α chain starts in the side wall of the dome, where it completes the fold of the MPT, then disappears in the interior of the reaction chamber and reappears ~60 Å away at the DM1 helices in the middle of the equatorial wheel. The sequence between these peripheral and central anchor points corresponds to the substrate-conjugating ACP domain, flanked by flexible linkers. This architectural solution implies that ACP is double-tethered diagonally across the reaction chamber (Fig. 4A).

We investigated the conformations of the ACP domains in the reaction chambers using cryoelectron microscopy. A single particle reconstruction of *T. lanuginosus* FAS at 18 Å resolution shows density features attributable to ACP domains inside the reaction chambers (Fig. 4C). The observed density does not allow fitting of the ACP fold (*14*), but rather reflects the time-averaged positions of ACP domains and the associated linkers in solution. The density extending from the central wheel can be attributed to an additional portion of the downstream ACP linker not observed in the crystallographic experiment.

The peripheral ACP anchor provides intriguing insights into the evolution of fungal FAS. The crystal structure reveals that the N terminus of the α chain is very close in space (~8 Å apart) to the location where the C terminus of the β chain ends (Fig. 4B). The two chains together describe the fold of the MPT, with the β chain contributing the portion with catalytic residues and the α chain completing the core fold of the MPT (fig. S9). In fact, certain species express fungal type I FAS from a single gene without the insertion of additional amino acids, when compared with the β- and α-chain termini of the *T. lanuginosus* or yeast FAS (fig. S6). The composite nature of the fungal MPT, which superimposes with the bacterial homolog (Fig. 4B) (*21*), indicates that the two genes encoding the α and β chains are unlikely to have evolved independently and that double-gene transcribed FASs are probably a result of gene splitting of a single ancestor. Single-gene FASs are found in *Ustilago maydis* and related fungi. Although, in different species, gene splitting occurs at various positions of the sequence (fig. S6), it is never observed within the ACP linkers, which is an argument for the existence of considerable evolutionary pressure in order to preserve two-sided covalent attachment of ACP in the reaction chambers.

Acetyltransferase and malonyl/palmitoyl transferase. The similar fold of AT and MPT reflects the related functions of these enzymes, namely, the transfer of acyl groups between CoA and ACP. Because of the position-specific intra- and intersubunit contacts of AT and MPT within the multienzyme complex, major structural differences are observed at the peripheries in the form of expansion segments (fig. S7). On the basis of biochemical and mutagenesis experiments performed with yeast FAS (*22–24*) and the comparison with bacterial malonyltransferase (MT) structures (*25, 26*), the catalytic histidines ($H^{440}_{AT(Tl)}/H^{2006}_{MPT(Tl)}$) and serines ($S^{286}_{AT(Tl)}/S^{1837}_{MPT(Tl)}$) could be assigned in the deep crevice between the large and small subdomains (Fig. 5A).

Previous studies of the yeast FAS transacylase specificities indicated that MPT preferentially transfers malonyl and palmitoyl moieties, whereas AT specifically reacts with the acetyl group (*4, 22–24*). The positively charged arginine $R^{1863}_{MPT(Tl)}$ located at the base of the catalytic pocket in MPT (Fig. 5A) probably recognizes the negatively charged free carboxyl group of the malonyl moiety, as observed in the *Escherichia coli* MT cocrystallized with malonyl-CoA (*26*). In AT, however, $R^{1863}_{MPT(Tl)}$ is replaced by the highly conserved isoleucine $I^{318}_{AT(Tl)}$, which should favor the binding of the uncharged acetyl moiety and should lead to discrimination against malonyl binding. Mutating the corresponding arginine residue in the bifunctional human malonyl/acetyltransferase significantly reduced malonyl and increased acetyltransferase activity (*27*), consistent with our structural observations.

Another important structural difference between AT and MPT is the narrowed catalytic cleft of AT, which might prevent binding of long-chain acyl substrates. In contrast, the catalytic cleft of MPT is open and contains a number of conserved hydrophobic residues well suited for binding the acyl chains of fully elongated and saturated C_{16} or C_{18} fatty acid products for back-transfer to CoA (figs. S8 and S9).

Ketoacyl synthase. The canonical KS dimers that catalyze the decarboxylative condensation of the acyl and malonyl substrates are tightly embedded in the central wheel of the FAS particle, and their catalytic clefts point into opposite reaction chambers (Fig. 5C). The broad specificity of fungal type I KS for C_2 to C_{16} acyl primer substrates contrasts with condensing enzymes of the dissociated bacterial type II FAS system, where acyl chains are elongated by three synthases (KASs) with different substrate specificities: KAS III (FabH) elongates C_2 by using acetyl-CoA as a primer, whereas KAS I (FabB) and KAS II (FabF) extend ACP-linked substrates with a primer length of C_4 to C_{14} and C_{16}, respectively (*1*). Structural superpositions

Fig. 4. Integration of the substrate-carrier domain ACP in fungal FAS. (**A**) Partially cut-away reaction chambers of FAS are shown from the side. Each of the two reaction chambers contain three double-tethered flexible ACP domains (schematically represented by cyan spheres and yellow dashed lines) between the peripheral anchors in the reaction chamber wall and central anchors at the base. (**B**) Structures of the central and peripheral ACP anchors shown in ribbon representation. The peripheral anchor, which corresponds to the N terminus of the α chain (blue), forms the MPT fold together with the C terminus of the β chain (green). For illustration, the bacterial MT (yellow) has been superimposed on the MPT. (**C**) Cryoelectron microscopic reconstruction of *T. lanuginosus* FAS at 18 Å resolution. The particle is sliced and additional density not observed in the electron density map of the crystal structure is colored red.
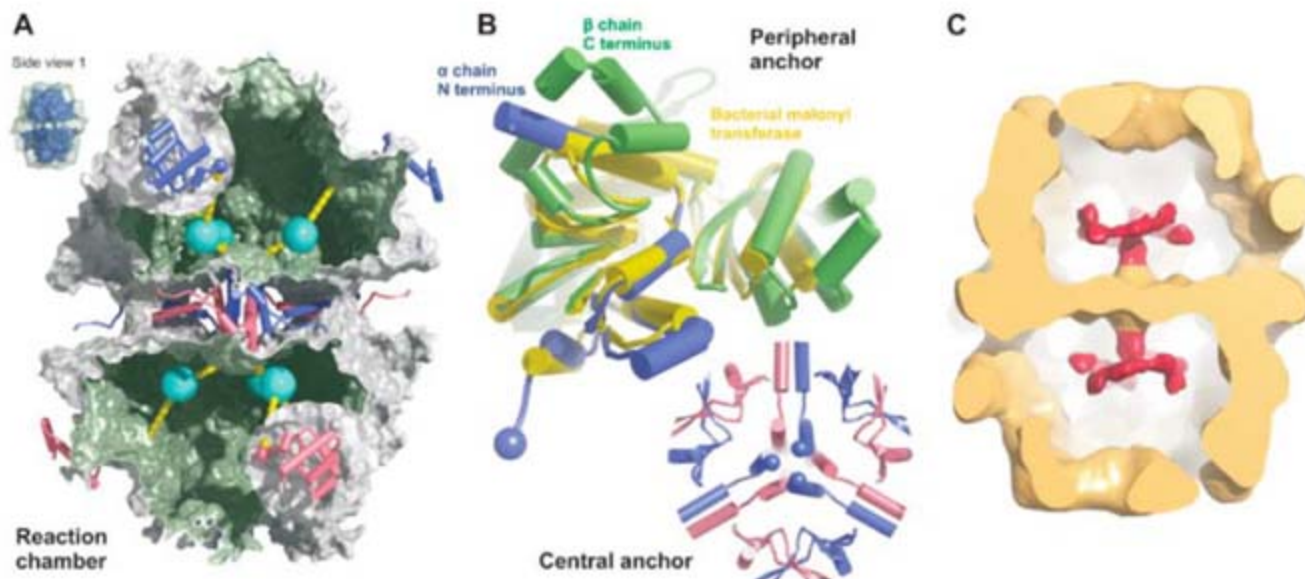
of the fungal KS thiolase fold with those of the three bacterial enzymes reveals a close resemblance to both KAS I and KAS II, whereas the structural relation to KAS III is more distant (fig. S7 and table S3). Note that the acyl-binding pocket of KS shares conserved features with both KAS I and KAS II, which may explain the broader specificity of the fungal KS (fig. S10). The use of a single KS for all condensation steps in the fungal FAS complex seems advantageous; however, it obviously necessitated the integration of an additional catalytic activity, the AT, to provide ACP-bound C_2 primer substrates.

Ketoacyl reductase. KR catalyzes the first of three β-carbon processing steps, the NADPH-dependent reduction of ketoacyl-ACP to hydroxyacyl-ACP. The catalytic core of fungal KR, which adopts a classical Rossmann fold, is highly similar to the bacterial type II enzymes (fig. S7 and table S3). Nevertheless, fungal KR forms dimers, in contrast to the tetrameric type II KRs (12, 28). Integrated into the fungal FAS complex, the active sites are located in opposite reaction chambers (Fig. 5B).

In the structure of *T. lanuginosus* FAS from crystals soaked with $NADP^+$, the cofactor binds in the active site cleft just above the equator of the central wheel (Fig. 5B and fig. S2). The reactive nicotinamide part of $NADP^+$ is directly accessible from the inside of the reaction chamber, and a long tunnel, which could accommodate the acyl chain, points away from the catalytic center (supporting online text and fig. S11). The comparison of the dinucleotide-free with the $NADP^+$-bound form of fungal KR reveals structural changes in the loop contacting the pyrophosphate moiety of $NADP^+$ (fig. S11). The loop adopts an "open" conformation in the empty KR form and closes upon cofactor binding. This conformational change might ensure that the NADPH cofactor, which is a prerequisite for catalysis, is bound to KR before the ketoacyl-ACP substrate. Such a scenario would be similar to type II KRs, where rearrangements of a corresponding shorter loop, as well as regions on the opposite side of the catalytic cleft have been observed after $NADP^+$ binding (28, 29).

Dehydratase. The elimination of water from hydroxyacyl-ACP is catalyzed by DH. The overall structure of fungal DH reveals a triple hot dog fold with the pseudodimeric part closely related to eukaryotic hydratase 2 enzymes, which catalyze the inverse reaction and use CoA-linked substrates (figs. S7, S13, and S14) (30–32). The additional hot dog domain is inserted between the two pseudosymmetric core subunits. On the basis of similarity with the hydratase 2 structures, the catalytic histidine $H^{1591}_{DH(Tl)}$ is located in the C-terminal domain of fungal DH, from which a ~20 Å–long hydrophobic tunnel runs into the N-terminal subdomain (supporting online text). The architecture of this substrate-binding tunnel illustrates potentially opposing requirements imposed onto all catalytic domains as they became incorporated into the multienzyme. DH has, for example, evolved insertions and an additional domain that stabilize the interactions between its catalytic core and the rest of the multienzyme. Nevertheless, the selective pressure has ensured that in spite of these additional features, the maximal length of the substrate-binding tunnel is maintained.

Enoyl reductase. The ER catalyzes the last step in the fatty acid elongation cycle by reducing the enoyl-ACP intermediate to saturated acyl-ACP. In contrast to most other FAS systems, where electrons from NAD(P)H are transferred directly to the substrate by a different catalytic mechanism, reduction by the fungal type I ER is flavin mononucleotide (FMN)–dependent (33). The fungal ER contains a triose phosphate isomerase (TIM)–barrel fold (unrelated to the Rossmann fold of the bacterial ER FabI) with a large, predominantly α-helical insertion (Fig. 5D). The FMN-binding pocket is located at the base of the TIM barrel, and $NADP^+$ binds in the deep cleft between the insertion domain and the barrel (fig. S16). A remarkably similar structural homolog (despite

Fig. 5. Active sites. (**A**) Superposition of the AT (green) and MPT (red) catalytic center. The arginine implicated in the recognition of the negatively charged malonyl moiety by forming a bidentate salt bridge in MPT is replaced by a hydrophobic isoleucine in the acetyl-specific AT. (**B**) Organization of the KR dimer in the central wheel of the FAS particle. Each KR dimer is tightly embedded into the central wheel. As a result, the two active sites of the dimer are located in different reaction chambers. The $NADP^+$ cofactor bound to the catalytic center of KR is shown in red. (**C**) Topology of KS in the FAS complex. As a consequence of the KS integration into the central wheel, the catalytic clefts of each KS dimer point into different reaction chambers. The deep, hydrophobic substrate-binding pockets extend to the dimer interface and are shown in red. (**D**) The FMN-containing fungal ER adopts a TIM barrel fold and is structurally unrelated to the mammalian and most bacterial counterparts. The peptide chain is colored rainbow from its N terminus (blue) to the C terminus (red). The $NADP^+$ cofactor (red) is bound in a deep cleft formed by the TIM barrel subdomain, which harbors FMN (yellow), and a large, mainly α-helical insertion. The $F_{obs} - F_{calc}$ omit Fourier map (gray) showing both cofactors is contoured at 3 σ. (**E**) The ER embedded into the wall of the FAS complex is colored according to its surface conservation from white (not conserved) to green (highly conserved). The NADPH cofactor can approach the binding cleft from the outside of the FAS particle. The highly conserved inner surface of ER allows direct access to the reactive FMN and the catalytic histidine.

Fig. 6. Substrate shuttling in the fungal FAS. (**A**) On the basis of the shortest distances between both ACP anchor points (green) and all catalytic domains in the reaction chamber, a complete set of catalytic sites necessary for the cyclic reaction can be defined (red). On account of the elasticity of linkers (yellow), free diffusion of ACP (cyan) would be "channeled" into a circular path, which further increases the local concentration of the ACP in the area where active sites are distributed (arrow). (**B**) The reaction chamber contains three complete sets of active sites, which are arranged in an approximate plane around the peripheral ACP anchors. Note that different α (pink, blue) and β chains (green, yellow, gray) contribute to each catalytic set.
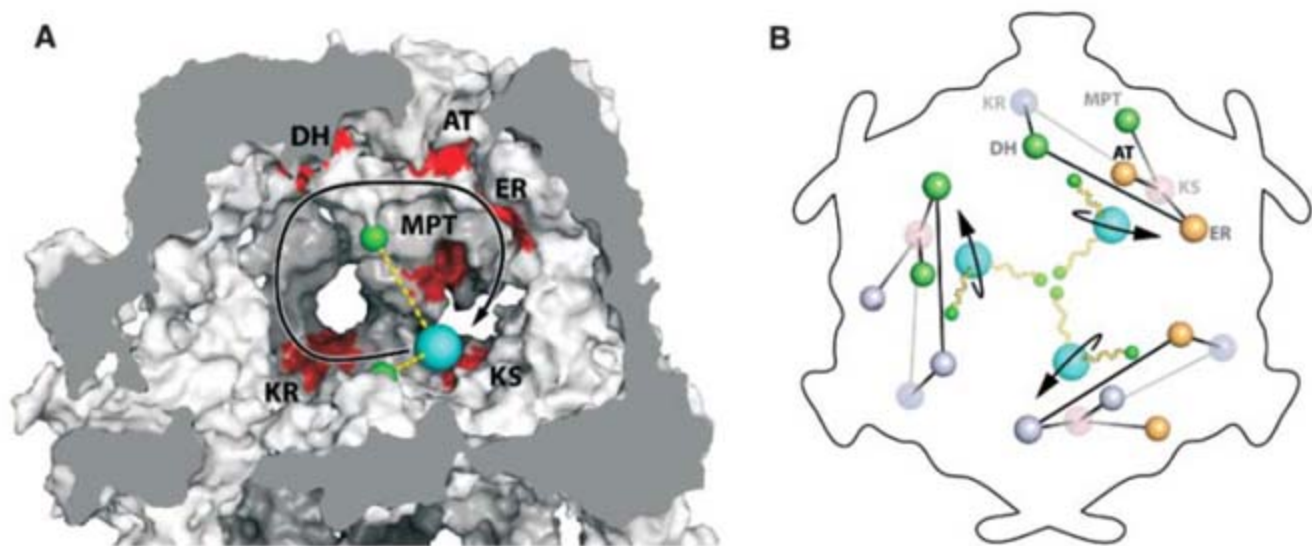
the low sequence homology of 16%) is the FMN-dependent 2-nitropropane dioxygenase (ND) from *Pseudomonas aeruginosa* (figs. S7 and S17 and table S3) (*34*). This suggests that the reduction step catalyzed by fungal ER may rely on analogous catalytic residues (supporting online text) as the oxidative denitrification of nitroalkenes catalyzed by ND.

The ER is optimally imbedded within the FAS complex to allow easy access of the NADPH to its binding pocket from the outside of the FAS particle, whereas its catalytic center is best accessible from the interior of the reaction chamber, where ACP delivers the substrates (Fig. 5E). The surface surrounding the active site harbors many conserved residues, which might represent a docking site for ACP (*14*). Superposition of our NADP$^+$-bound ER structure with other FMN-containing TIM barrel oxidoreductases in complex with their substrates (*34*, *35*) reveals that the substrates would sterically clash with the nicotinamide moiety of NADP$^+$, which suggests that oxidized NADP$^+$ is likely released before enoyl-substrate binding. Such sequential and mutually exclusive substrate binding would imply a two-step Ping-Pong catalytic mechanism for ER of type I fungal FAS.

The structure of the fungal ER also rationalizes the observation that some pathogenic bacteria, such as *Streptococcus pneumoniae*, display resistance to triclosan, a well-known antibiotic and inhibitor of the common FabI-like ER (*36*, *37*). These resistant bacteria lack the *fabI* gene present in triclosan-sensitive bacteria, but instead have a different essential gene, *fabK*. Interestingly, an alignment of FabK sequence and fungal ER identifies FabK as a TIM-barrel–containing flavoenzyme similar to fungal ER, having conserved FMN-interacting residues and the same proposed catalytic histidine (fig. S17). Therefore, these bacteria probably acquired an unusual, fungal-

related TIM barrel fold ER that is not inhibited by triclosan.

Topology of catalytic centers in the FAS assembly. For efficient catalysis by FAS, short diffusion distances for ACP during substrate transfer are crucial, as well as unhindered access to active site clefts. In order to define the likely path of an individual ACP during the reaction cycle, we measured the distances between all catalytic sites in one reaction chamber and the two anchor points for one ACP domain. Although the maximal extended length of the flexible linkers would allow ACP to reach many catalytic sites within the reaction chamber (*38*), one comprehensive set of catalytic sites is located considerably closer to an individual set of anchor points than any other combination of active sites (table S6). Moreover, the catalytic sites of this set are distributed in a circular manner around the peripheral, elevated anchor point, and the substrate clefts of the enzymes point toward the anchors (Fig. 6A).

This proposed reaction pathway allows ACP, after substrate loading by AT or MPT, to reach all subsequent catalytic centers in a clockwise motion when viewed from the inside of the reaction chamber, mirroring the usual textbook depiction of the fatty acid synthesis cycle. Consequently, after one full reaction cycle, the acyl chain is back-transferred from ACP to the same KS for the next elongation step, and the growing fatty acid is not handed over from one ACP to another.

The distances between the two ACP anchor points and the respective set of active sites is relatively uniform, ~85 ± 15 Å for the anchor point within the central wheel and ~55 ± 15 Å for the anchor within the MPT (Fig. 6A and table S6). Nuclear magnetic resonance experiments have shown that alanine- and proline-rich linker sequences have some degree of rigidity in solution (*39*, *40*). Single attachment of ACP would allow it to move within a sphere defined

by the persistence length of the linker. However, the observed double attachment of ACP favors its movement along the intersection of the two spheres, which define a "low-energy circle" perpendicular to the vector that connects the anchor points. The length and the sequence of the linkers would affect the steepness of the three-dimensional energetic landscape around this circle and must have evolved to match the approximately circular distribution of the active sites. A mechanical analogy to this mechanism would be a ball connected to separate anchor points via two differently sized springs whose length in relaxed state is greater than the distance between the two anchors. Without stretching or compressing the springs, such a ball could only pivot around the axis that connects the two anchor points.

It is evident from the topology of the α subunit in the FAS complex that for each α chain, the KS catalytic cleft points into one, and the KR active site into the other, reaction chamber. Similarly, KS and the ACP anchors of the same subunit are located in different reaction compartments. Therefore, at least two α chains are required for catalysis (Fig. 6B). This is in agreement with early cross-linking experiments (*41*), showing that dibromopropanone cross-linking of the PPT group of ACP with the active site cysteine of KS resulted in oligomerization of the FAS α chains. Inspection of the proposed pathway for ACP during a full reaction cycle shows that it encounters catalytic enzymes that are distributed not only across two α chains, but also over two β chains (Fig. 6B). If only a single β subunit was involved in the reaction cycle, the overall travel distance of ACP would be considerably larger.

The first high-resolution structures of type I fungal FAS reveal the molecular basis of compartmentalized fatty acid synthesis in a complex, multienzymatic megasynthase. These results now make it possible to evaluate this interde-

pendent system of enzymatic reactions in a well-defined structural framework with a range of features that are not considered in single-enzyme catalysis. Moreover, detailed mechanistic questions regarding fatty acid synthesis by fungal FAS and related type I FAS enzymes (supporting online text) can now be addressed with structure-based genetic and biochemical experiments.

References and Notes

1. S. W. White, J. Zheng, Y. M. Zhang, C. O. Rock, *Annu. Rev. Biochem.* **74**, 791 (2005).
2. J. W. Campbell, J. E. Cronan, *Annu. Rev. Microbiol.* **55**, 305 (2001).
3. S. J. Wakil, J. K. Stoops, V. C. Joshi, *Annu. Rev. Biochem.* **52**, 537 (1983).
4. E. Schweizer, J. Hofmann, *Microbiol. Mol. Biol. Rev.* **68**, 501 (2004).
5. S. Smith, A. Witkowski, A. K. Joshi, *Prog. Lipid Res.* **42**, 289 (2003).
6. F. Lynen, *Eur. J. Biochem.* **112**, 431 (1980).
7. F. J. Asturias *et al.*, *Nat. Struct. Mol. Biol.* **12**, 225 (2005).
8. T. Maier, S. Jenni, N. Ban, *Science* **311**, 1258 (2006).
9. F. Wieland, E. A. Siess, L. Renner, C. Verfürth, F. Lynen, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5792 (1978).
10. J. K. Stoops, S. J. Kolodziej, J. P. Schroeter, J. P. Bretaudiere, S. J. Wakil, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6585 (1992).
11. S. J. Kolodziej, P. A. Penczek, J. K. Stoops, *J. Struct. Biol.* **120**, 158 (1997).
12. S. Jenni, M. Leibundgut, T. Maier, N. Ban, *Science* **311**, 1263 (2006).
13. Materials and methods are described in the supporting material on *Science* Online.
14. M. Leibundgut, S. Jenni, C. Frick, N. Ban, *Science* **316**, XXX (2006).
15. M. Ishikawa, D. Tsuchiya, T. Oyama, Y. Tsunaka, K. Morikawa, *EMBO J.* **23**, 2745 (2004).
16. R. N. Perham, *Annu. Rev. Biochem.* **69**, 961 (2000).
17. J. K. Stoops *et al.*, *J. Biol. Chem.* **253**, 4464 (1978).
18. F. Lynen, *Biochem. J.* **102**, 381 (1967).
19. A. M. Haapalainen, G. Merilainen, R. K. Wierenga, *Trends Biochem. Sci.* **31**, 64 (2006).
20. Y. Tang, C. Y. Kim, I. I. Mathews, D. E. Cane, C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11124 (2006).
21. L. Serre, E. C. Verbree, Z. Dauter, A. R. Stuitje, Z. S. Derewenda, *J. Biol. Chem.* **270**, 12961 (1995).
22. H. Engeser, K. Hübner, J. Straub, F. Lynen, *Eur. J. Biochem.* **101**, 407 (1979).
23. H. Engeser, K. Hübner, J. Straub, F. Lynen, *Eur. J. Biochem.* **101**, 413 (1979).
24. H. Schuster, B. Rautenstrauss, M. Mittag, D. Stratmann, E. Schweizer, *Eur. J. Biochem.* **228**, 417 (1995).
25. A. T. Keatinge-Clay *et al.*, *Structure* **11**, 147 (2003).
26. C. Oefner, H. Schulz, A. D'Arcy, G. E. Dale, *Acta Crystallogr. D Biol. Crystallogr.* **62**, 613 (2006).
27. V. S. Rangan, S. Smith, *J. Biol. Chem.* **272**, 11975 (1997).
28. A. C. Price, Y. M. Zhang, C. O. Rock, S. W. White, *Structure* **12**, 417 (2004).
29. A. C. Price, Y. M. Zhang, C. O. Rock, S. W. White, *Biochemistry* **40**, 12772 (2001).
30. Y. M. Qin *et al.*, *J. Biol. Chem.* **275**, 4965 (2000).
31. M. K. Koski, A. M. Haapalainen, J. K. Hiltunen, T. Glumoff, *J. Biol. Chem.* **279**, 24666 (2004).
32. K. M. Koski, A. M. Haapalainen, J. K. Hiltunen, T. Glumoff, *J. Mol. Biol.* **345**, 1157 (2005).
33. J. L. Fox, F. Lynen, *Eur. J. Biochem.* **109**, 417 (1980).
34. J. Y. Ha *et al.*, *J. Biol. Chem.* **281**, 18660 (2006).
35. P. A. Hubbard, X. Liang, H. Schulz, J. J. Kim, *J. Biol. Chem.* **278**, 37553 (2003).
36. R. J. Heath, C. O. Rock, *Nature* **406**, 145 (2000).
37. H. Marrakchi *et al.*, *Biochem. J.* **370**, 1055 (2003).
38. K. Werkmeister, R. B. Johnston, E. Schweizer, *Eur. J. Biochem.* **116**, 303 (1981).
39. S. E. Radford, E. D. Laue, R. N. Perham, S. R. Martin, E. Appella, *J. Biol. Chem.* **264**, 767 (1989).
40. R. N. Perham, *Biochemistry* **30**, 8501 (1991).
41. J. K. Stoops, S. J. Wakil, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4544 (1980).
42. All data were collected at the Swiss Light Source (SLS, Paul Scherrer Institute, Villigen). We are grateful to C. Schulze-Briese, S. Gutmann, E. Pohl, S. Russo, and T. Tomizaki for their outstanding support at the SLS. We thank T. Maier and J. Erzberger for critically reading the manuscript and all members of the Ban laboratory for suggestions and discussions; R. Grosse-Kunstleve, P. Afonine, and P. Adams for providing a prerelease version of the PHENIX refinement program and advice regarding structure refinement; A. Jones for a prerelease version of the program O; and D. Sargent for technical assistance. We thank A. Tsang for assembling *T. lanuginsosus* FAS contigs from expressed sequence tag sequence data produced by the Fungal Genomics Project (https://fungalgenomics.concordia.ca). This work was supported by the Swiss National Science Foundation (SNSF) and the National Center of Excellence in Research (NCCR) Structural Biology program of the SNSF. D.B. was supported by a Federation of European Biochemical Societies long-term fellowship. We thank the Electron Microscopy Center Zurich (EMEZ) for support. Coordinates and structure factors of the *T. lanuginosus* crystal structures have been deposited in the Protein Data Bank with accession codes 2UV9, 2UVA, 2UVB, and 2UVC. The cryoelectron microscopy map has been deposited at the EM Data Bank with accession code EMD-1338.

REPORTS

Plastic and Moldable Metals by Self-Assembly of Sticky Nanoparticle Aggregates

Rafal Klajn,[1]* Kyle J. M. Bishop,[1]* Marcin Fialkowski,[1] Maciej Paszewski,[1] Christopher J. Campbell,[1] Timothy P. Gray,[2] Bartosz A. Grzybowski[1,2]†

Deformable, spherical aggregates of metal nanoparticles connected by long-chain dithiol ligands self-assemble into nanostructured materials of macroscopic dimensions. These materials are plastic and moldable against arbitrarily shaped masters and can be thermally hardened into polycrystalline metal structures of controllable porosity. In addition, in both plastic and hardened states, the assemblies are electrically conductive and exhibit Ohmic characteristics down to ~20 volts per meter. The self-assembly method leading to such materials is applicable both to pure metals and to bimetallic structures of various elemental compositions.

Although various techniques have been developed to assemble nanoparticles at the microscale, such as nanoparticle superlattices (*1*, *2*) and three-dimensional crystals (*3*, *4*), scaling up these procedures to freestanding, macroscopic materials (*5*) has proven challenging. This situation contrasts with the formation of ceramic "greenwares" from colloidal particles such as clays, the interactions of which can hold the particles close together even during the dehydration and calcining steps that produce the final product (*6*). Here, we describe straightforward synthesis of such materials by a two-step method in which individual metal nanoparticles first self-assemble into deformable spherical aggregates ("supraspheres") and then "glue" together like pieces of clay into millimeter-sized structures. Notably, not only are these macroscopic materials plastic and moldable into arbitrary shapes, but they also can be electrically conductive. In addition, because of the metastability of the spherical building blocks, they can be thermally hardened and structurally evolved into porous, polycrystalline metal monoliths. The described methodology works well with nanoparticles of more than one type and can be used to prepare both pure-metal and bimetallic materials.

We used spherical nanoparticle assemblies, or supraspheres, similar to those prepared previous-

[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA. [2]Department of Chemistry, Northwestern University, Evanston, IL 60208, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: grzybor@northwestern.edu
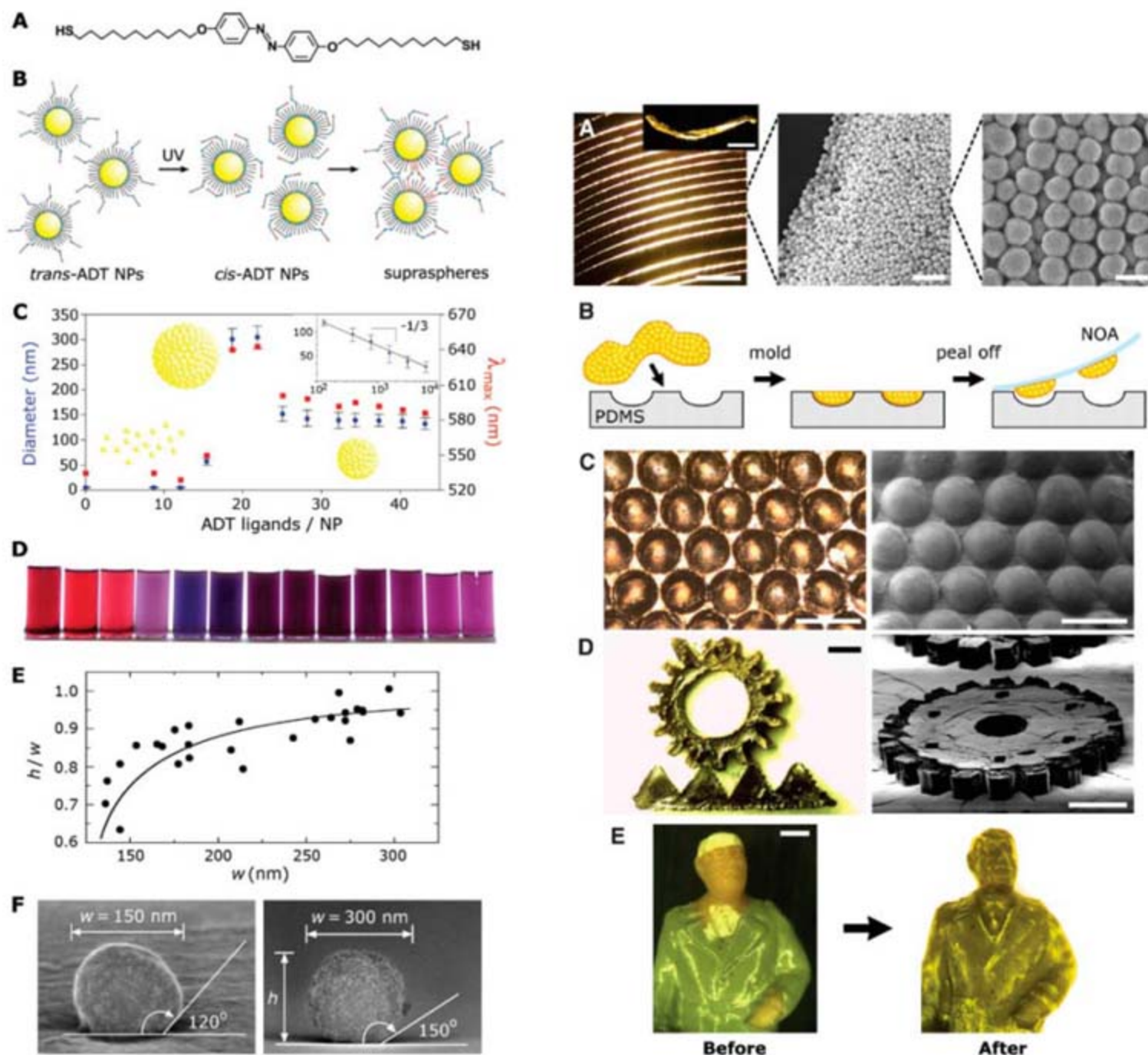
Fig. 1. Synthesis and properties of spherical nanoparticle aggregates. (**A**) Structure of the dithiol azobenzene cross-linker [4,4'-bis(11-mercaptoun-decanoxy)azobenzene (ADT)] mediating self-assembly. (**B**) Scheme of the light-induced self-assembly of metal nanoparticles (NPs) coated with DDA and ADT into supraspheres (for mechanistic details of this process, see SOM text, sections 2 and 3). (**C**) Diameters (blue) and the wavelengths of maximal absorption (λ_{max}, red) for supraspheres formed by UV irradiation of Au nanoparticles (yellow) covered with various numbers of ADT ligands. Below the critical number of ligands adsorbed onto each nanoparticle (here, ~20), the photoinduced interactions between nanoparticles are too weak to cause their aggregation. Above this limit, nucleation and growth of supraspheres occurs. (Inset) Suprasphere diameters, D, decrease with increasing concentration of ADT ligands. The abscissa has the total ADT concentration (both adsorbed and free ligands) and is logarithmic to emphasize the agreement of the experimental results with $D \sim C_{ADT}^{-1/3}$ power law predicted by the nucleation-and-growth model (black line). Error bars show mean ± SD. (**D**) For metals exhibiting surface plasmon resonance (Au and Ag), aggregation is accompanied by a pronounced color change (as shown here for Au). The vials in the bottom row are solutions of supraspheres arranged in the order of increasing ADT/NP ratio and correspond to the points in the plot.

(**E**) Aspect ratios of height to width, h/w, of supraspheres placed on silicon (100) increase with width. Solid line corresponds to the trend predicted by the Ramberg-Osgood model of suprasphere plasticity (for more details, see SOM text, section 4). (**F**) Side-view SEM images of two typical supraspheres on a silicon surface. **Fig. 2.** (**A**) Upon dewetting of a suprasphere solution from an inclined silicon surface, the supraspheres—similar to coffee dregs (24)—organize into bands. (Left) An optical micrograph of such bands ("wires") while still attached to the surface. Scale bar, 1 mm. (Inset) A freestanding wire washed off the surface with methanol. Scale bar, 0.5 mm. (Middle) SEM image showing the wire's porous structure. Scale bar, 1 μm. (Right) A magnification of individual supraspheres. Scale bar, 200 nm. (**B**) Concentrated solution of the supraspheres can be molded against masters [here, made of polydimethylsiloxane (PDMS)] of arbitrary surface reliefs and then peeled off onto a polyurethane film [Norland Optical Adhesive (NOA) no. 68]. (**C** and **D**) Optical and SEM images of microstructures prepared by molding Au or Ag supraspheres against an array of concave microlenses and an array of microgears. The freestanding gear and track pieces in the left panel of (D) were arranged manually after liberating from the polyurethane film. Scale bars, 200 μm. (**E**) Polypropylene miniature before (left) and after (right) coating with supraspheres. Scale bar, 1 mm.

ly (7–9). For our purposes, however, we required that the supraspheres be structurally sturdy as well as deformable and adhesive with respect to one another. We expected that these proper-

ties could be achieved by using long-chain dithiol cross-linkers (Fig. 1A), the terminal thiol (SH) groups of which are known to bind strongly to noble metal nanoparticles. The long

spacer connecting these groups should endow the resulting aggregates with flexibility. In addition, our linkers incorporated a photo-switchable azobenzene unit that allowed for precise control of the assembly process and of the dimensions of the forming spheres by ultra-violet (UV) light.

In a typical procedure, we used 2 mM solutions of metal nanoparticles (here, mostly Au, but also Ag, Pt, or Pd) in toluene. The nanoparticles were ~5 nm in diameter and had a dispersity of $\sigma \approx 10\%$ [supporting online material (SOM) text, section 1]. We stabilized the solutions with 35 mM dodecylamine (DDA) capping agent and 10 mM didodecyldimethyl-ammonium bromide (DDAB) surfactant. Next, under vigorous stirring, we added varying amounts (up to 2.4 mM) of photoactive trans-azobenzene dithiol (ADT) ligands (Fig. 1B) to the solutions. Because excess surfactant and capping agent prevented the spontaneous cross-linking of nearby nanoparticles through their divalent ADT ligands, the unirradiated solutions were stable (10) for many weeks. Low-power UV irradiation (365 nm, 0.7 mW/cm²) (11) caused rapid trans-cis isomerization of the ADTs (Fig. 1B, center) and induced mo-lecular dipoles on the azobenzene units [electric dipole moment (μ) = 4.4 debye for the cis (12) form compared with 0 debye for trans], causing the molecules to aggregate and cross-link into supraspheres (Fig. 1B, right). [For detailed discussion of the forces mediating self-assembly, see SOM text, sections 2 and 3.] The suprasphere growth occurred by means of a nucleation-and-growth mechanism, in which the free nanoparticles initially nucleated into small, thermodynamically stable clusters (un-less smaller than a critical size) that sub-sequently grew by the addition of single nanoparticles until all nanoparticles available were used (13). The important feature of this mechanism is that it allowed the control of suprasphere diameters (Fig. 1C) by changing the concentration of the nanoparticles and/or the concentration of the ADT cross-linkers (14, 15). In this way, we prepared supraspheres with diameters between about 50 and 300 nm (16) (Fig. 1, C and D, and fig. S6).

Irrespective of their sizes, the supraspheres were highly deformable upon contact with other spheres (or surfaces) and—despite being com-posed in ~92% weight by weight of metal (17)—had the mechanical properties of a plastic solid. The plasticity was confirmed in two ways: (i) by nanoindentation of individual supraspheres and (ii) by the aspect ratios and contact angles of spheres adhered onto different substrates, always increasing with increasing sphere size. Here, a systematic study of supra-spheres on silicon (100) surfaces indicated (Fig. 1, E and F) that their deformation is driven by short-ranged adhesion forces, and can be described by the so-called Ramberg-Osgood model (18) of initial elastic response followed

Fig. 3. (A) Typical load-displacement loops recorded during nanoindentation (TriboIndenter, Hysitron Inc.) of (left) a plastic wire (~1 mm by 170 μm by 30 μm) made of ~150-nm gold supraspheres and (right) the same wire after hardening at 50°C for 12 hours. The compression-release hysteresis recorded before hardening indicates plastic deforma-tion. (B) SEM images corresponding to the graphs above show structural changes accompanying hardening. (C) Hardened materials subject to further heating at 350°C melt and shrink to ultimately give a solid block of metal. The panels show the progression, from top left, to bottom left, to bottom right, and finally to top right. Scale bars, 200 nm.
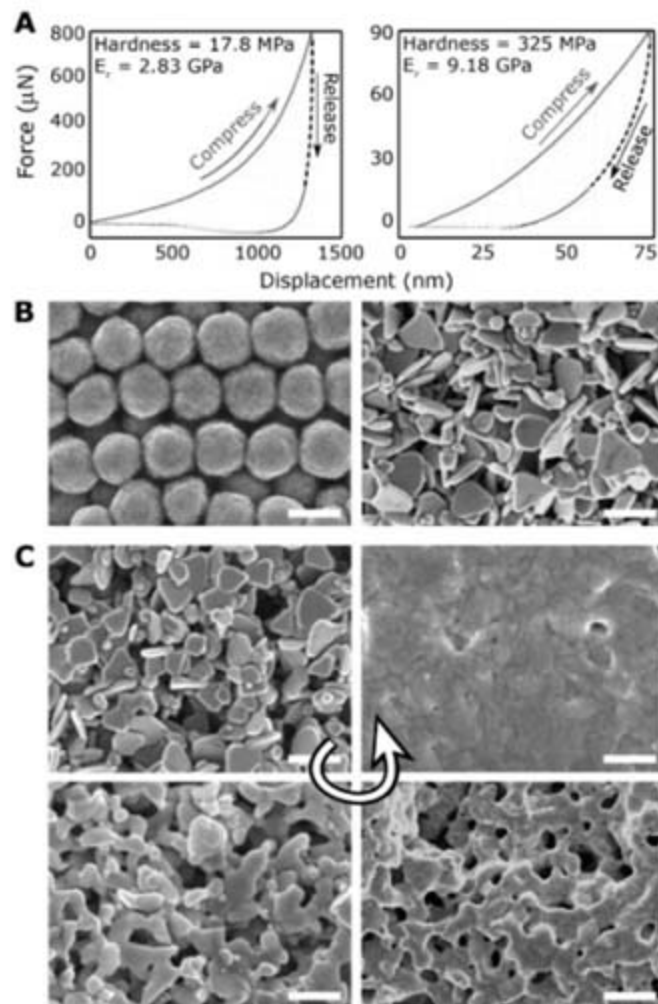
Fig. 4. Assembly schemes, SEM images, and energy dispersive analysis by x-rays (EDX) (Hitachi S3500) spectra of nanoporous materials made of (A) pure plati-num, (B) 1:1 gold-palladium, and (C) 3:1 gold-silver. To verify spa-tially uniform composition of the bimetallic materials, EDX spectra were taken by focusing the beam at different regions of the same sam-ple. a.u., arbitrary units. Scale bars, 500 nm.
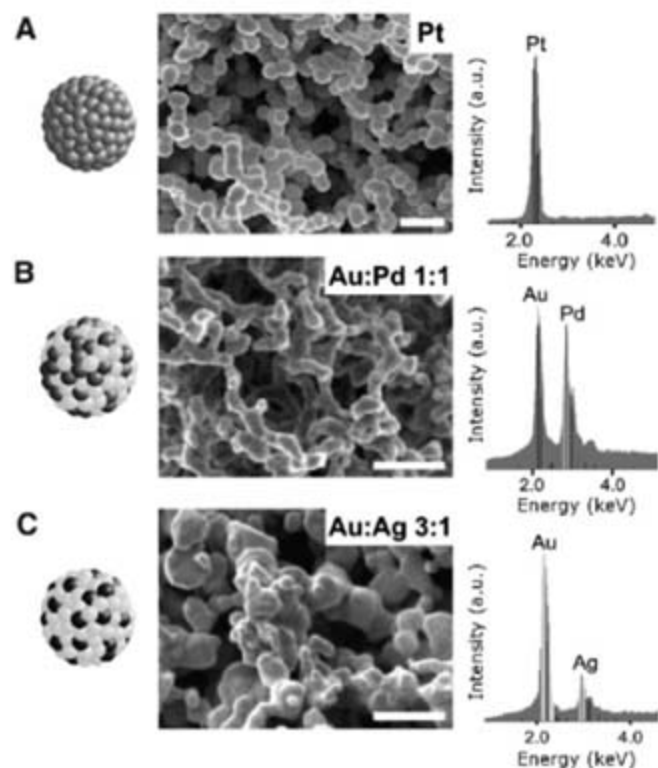
by irreversible, plastic deformation (SOM text, section 4).

The plasticity and "stickiness" of the supraspheres enabled their further assembly into macroscopic materials (19). This is shown in Fig. 2, where suprasphere solutions either were dewetted from an inclined surface to leave behind an array of parallel "wires" (Fig. 2A) or were first concentrated to between 0.1 and 0.5% of the original volume and then molded against arbitrarily shaped and patterned masters (Fig. 2, B to E). Upon drying the solvent and washing with methanol to remove excess surfactant, the structures obtained could be gently pealed off the surface onto a polyurethane resin and then mechanically detached to give freestanding materials. Nanoindentation experiments revealed that similar to the supraspheres they comprise, these materials were plastic (with reduced modulus E_r of few gigapascals and hardness H of a few tens of megapascals; Fig. 3, A and B, left panels). At the same time, they could be structurally evolved and hardened by mild thermal treatment (about 50°C) to give relatively brittle, polycrystalline porous materials ($E_r \sim 10$ GPa and $H \sim 300$ MPa; Fig. 3, A and B, right panels). The plastic-to-porous metal evolution proceeded without noticeable changes in the materials' overall dimensions and was the result of gradual desorption of DDA and dithiol molecules (20, 21) and consequent fusion of nearby nanoparticles composing the materials.

Because this process was similar for plastic precursors assembled from supraspheres of different sizes, it was possible to prepare hardened materials of different porosities. Scanning electron microscopy (SEM) analysis showed that the pore sizes were commensurate and increasing with increasing diameters, D, of the suprasphere building blocks. At the same time, surface areas of the porous materials decreased slightly with increasing D and, as evidenced by Brunauer-Emmett-Teller measurements, were typically several square meters per gram (e.g., 8 m^2/g for porous Ag made of 80-nm Ag supraspheres).

When hardening was continued at temperatures above 300°C, the pore sizes and the material's overall dimensions gradually decreased (by as much as 20%) to ultimately give a block of solid metal (Fig. 3C). During such heating, the material initially preserved its shape (up to about 10% shrinkage), but its fine structural details were lost upon full melting. Because of the adhesiveness of the supraspheres, they not only assembled with one another but also readily coated various glasses and plastics (e.g., rubber and polyurethanes such as Norland Optical Adhesive, SU-8 photoresist) dipped into suprasphere solutions. In one example of this capability (Fig. 2E), gold supraspheres formed uniform coating on a small polypropylene statuette.

In addition to plasticity, moldability, and susceptibility to thermal hardening, our evolving nanostructured materials exhibited electric properties reminiscent of semimetals. In both plastic and hardened states, these materials were electrically conductive; they had purely Ohmic current density-field strength characteristics over a range of applied electric fields from 2×10^5 V/m down to 20 V/m; and their resistance decreased as the underlying structure evolved (SOM text). The observed Ohmic characteristics are in sharp contrast to those of other metal-insulator structures studied to date, which usually require a threshold voltage (22, 23) for conductance.

Because our self-assembly approach allows for the use of different types of supraspherical building blocks, all of the discussed methods and material properties could be extended to nanostructured materials composed of more than one metal—that is, to bimetallic materials (Fig. 4). In principle, such materials could be made either from mixtures of supraspheres of different types (e.g., all Au and all Pt) or from mixed supraspheres composed of different types of nanoparticles. In practice, however, the first method proved unsatisfactory because the different metals often phase separated upon thermal hardening. In contrast, hardening of mixed supraspheres gave structurally uniform materials (on the scale of ~100 nm), in which the composition at every probed location was equal to that of the supraspheres used. In this way, we were able to prepare bimetallic monoliths of various combinations and proportions of Au, Ag, Pt, and Pd. Two representative examples—Au and Pd (1:1) and Au and Ag (3:1)—are shown in Fig. 4, B and C. We believe that because of their high porosity, these materials may find uses in separations science and in heterogeneous catalysis, in which the application of multicomponent "nanofoams" to catalyze multiple (or multistep) reactions seems particularly interesting.

References and Notes

1. C. J. Kiely et al., Nature 396, 444 (1998).
2. E. V. Shevchenko et al., Nature 439, 55 (2006).
3. A. M. Kalsin et al., Science 312, 420 (2006).
4. C. B. Murray, C. R. Kagan, M. G. Bawendi, Science 270, 1335 (1995).
5. F. Stellacci, Nat. Mater. 4, 113 (2005).
6. W. D. Kingery, H. K. Bowen, D. R. Uhlmann, Introduction to Ceramics (Wiley, New York, 1976).
7. A. K. Boal et al., Nature 404, 746 (2000).
8. I. Hussain, Z. X. Wang, A. I. Cooper, M. Brust, Langmuir 22, 2938 (2006).
9. M. M. Maye et al., J. Am. Chem. Soc. 127, 1519 (2005).
10. Excess DDA and DDAB helped to minimize van der Waals (vdW) forces between the nanoparticles by reducing the dielectric contrast ($\Delta\varepsilon$) between the nanoparticles and the solvent. Without this stabilizing effect, the magnitude of vdW interactions could be greater than k_BT, where k_B is the Boltzmann constant and T is temperature (e.g., with a Hamaker constant $A = 0.4 \times 10^{-21}$ J) for interacting alkane-chain molecules in toluene, $E_{vdW} \sim 2k_BT$), and the
11. Low-power irradiation for periods of up to several tens of minutes did not cause desorption of the thiols from nanoparticle surfaces. With more intense sources of UV light (>50 mW/cm^2) or at elevated temperatures, however, desorption was observed, and the nanoparticles coalesced into larger particles and aggregates that precipitated from solution.
12. X. Tong, G. Wang, A. Soldera, Y. Zhao, J. Phys. Chem. B 109, 20281 (2005).
13. The nucleation-and-growth mechanism applies to this system only at short time scales, during which the supraspheres assemble by means of dipole-dipole interactions to consume all free nanoparticles in solution. At longer time scales, the suprasphere aggregates become "cemented" by dithiol cross-links and cannot coalesce with one another to achieve full phase separation.
14. Because the number of nucleation sites is proportional to the concentration of ADT, the suprasphere diameter scales with the concentration of ADT as $D \sim C_{ADT}^{-1/3}$.
15. D. Kashchiev, Nucleation: Basic Theory with Applications (Butterworth-Heinemann, New York, 2000).
16. Light-controlled self-assembly gave supraspheres of smaller size dispersity than alternative methods described in the literature (7–9). Also, we verified that spheres connected by shorter dithiol ligands were less "sticky" and required prolonged heating at high temperatures in order to self-assemble into larger structures. These properties made such supraspheres unsuitable for the molding experiments we conducted.
17. Metal weight percent was estimated from that of underlying gold nanoparticles. Specifically, for a spherical core-shell nanoparticle with a core of radius $R_c \sim 2.8$ nm and density $\rho_c = 19.3$ g/cm^3 (for gold) and with a self-assembled monolayer (SAM) of thickness ~1.4 nm and density $\rho_{SAM} \sim 0.73$ (for linear hydrocarbons), the weight percentage of metal may be estimated as $\rho_c R_c^3 / [\rho_c R_c^3 + \rho_{SAM}(R_{SAM}^3 - R_c^3)] \sim 0.92$. Similar values are expected for other noble metals.
18. H. W. Haslach Jr., R. W. Armstrong, Deformable Bodies and Their Material Behavior (Wiley, New York, 2004).
19. When nanoparticles were simply cross-linked without forming the intermediate supraspheres, they formed precipitates that could not be assembled, molded, or hardened.
20. D. V. Leff, P. C. Ohara, J. R. Heath, W. M. Gelbart, J. Phys. Chem. 99, 7036 (1995).
21. H. M. Schessler, D. S. Karpovich, G. J. Blanchard, J. Am. Chem. Soc. 118, 9645 (1996).
22. For metal/insulator structures of dimensions comparable to our supraspheres, the threshold voltages reported were on the order of ~2 × 10^6 V/m [e.g., (23)].
23. A. Beck et al., Appl. Phys. Lett. 77, 139 (2000).
24. R. D. Deegan et al., Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 62, 756 (2000).
25. We thank M. Ratner and G. C. Schatz for helpful discussions, as well as S. Smoukov for his help with nanoindentation experiments. This work was supported by the NSF CAREER (CTS-0547633) Award, 3M Non-Tenured Faculty Award, the Pew Scholarship, and the Sloan Fellowship (to B.A.G.). R.K. was supported by the NSF under the Northwestern Materials Research Science and Engineering Center Award. K.J.M.B. was supported by the NSF Graduate Fellowship. C.J.C. was supported by a Northwestern University Presidential Fellowship.

Optimizing the Laser-Pulse Configuration for Coherent Raman Spectroscopy

Dmitry Pestov,[1]* Robert K. Murawski,[1,2] Gombojav O. Ariunbold,[1] Xi Wang,[1]
Miaochan Zhi,[1] Alexei V. Sokolov,[1] Vladimir A. Sautenkov,[1] Yuri V. Rostovtsev,[1,2]
Arthur Dogariu,[2] Yu Huang,[2] Marlan O. Scully[1,2]

We introduce a hybrid technique that combines the robustness of frequency-resolved coherent anti-Stokes Raman scattering (CARS) with the advantages of time-resolved CARS spectroscopy. Instantaneous coherent broadband excitation of several characteristic molecular vibrations and the subsequent probing of these vibrations by an optimally shaped time-delayed narrowband laser pulse help to suppress the nonresonant background and to retrieve the species-specific signal. We used this technique for coherent Raman spectroscopy of sodium dipicolinate powder, which is similar to calcium dipicolinate (a marker molecule for bacterial endospores, such as *Bacillus subtilis* and *Bacillus anthracis*), and we demonstrated a rapid and highly specific detection scheme that works even in the presence of multiple scattering.

The Raman vibrational spectrum of molecules provides an excellent fingerprint for species identification. Because of the Raman effect, lower-frequency (Stokes) radiation is emitted (Fig. 1A) when light irradiates a molecule. The signal is weak, but with the advent of powerful lasers, spontaneous Raman spectroscopy became a useful technique.

When the molecules are put into coherent oscillation by a pair of preparation pulses (pulses 1 and 2 in Fig. 1B) and a third pulse is scattered off of this coherent molecular vibration, a strong anti-Stokes signal (pulse 4) is generated (Fig. 1B). This process is called coherent anti-Stokes Raman scattering (CARS) [(*1–3*) and references therein].

Unfortunately, CARS from the molecules of interest is frequently obscured by the nonresonant (NR) four-wave mixing (FWM) signal from other molecules (Fig. 1C), or even from the same molecules, because of contributions from multiple off-resonant vibrational modes and the instantaneous electronic response. This unwanted NR FWM is often much stronger than the resonant signal because there are usually many more background molecules than target molecules. Fluctuations of the NR background can, and frequently do, completely wash out the CARS signature (*4*). A variety of methods, including polarization-sensitive techniques (*5*) and heterodyne (*6, 7*) and interferometric (*8–11*) schemes, have been developed to increase the signal-to-background ratio. However, these methods do not work well in the presence of strong multiple scattering in rough samples because scattering randomizes

spectral phases and polarization. For this reason, applications of CARS for rapid detection and recognition of strongly scattering media, such as anthrax spores, have often been deemed impractical.

In our recent work, we have developed techniques for maximizing the coherent molecular oscillation and minimizing the NR background. A sequence of femtosecond pulses is used so that pulses 1 and 2 prepare a coherent molecular vibration (corresponding to a particular mode of oscillation) and a third time-delayed probe pulse is scattered off of the oscillation, yielding the anti-Stokes signal. By delaying the third pulse relative to the first two, the NR FWM signal is eliminated. The combination of shaped preparation pulses and an ultrashort time-delayed probe pulse maximizes the signal and lessens the background contribution. We call this approach femtosecond adaptive spectroscopic technique via CARS (FAST-CARS) (*12*). The definitive fingerprint information is retrieved from the probe-delay dependence, but the delay tuning slows down the acquisition and makes the technique vulnerable to fluctuations.

In the present work, we used ultrashort (~50 fs) transform-limited pulses 1 and 2 for uniform broadband Raman excitation and a tailored narrowband (several hundred femtosecond long) probe. We used a multichannel detector to simultaneously record the anti-Stokes signal at all optical frequencies within the band of interest. These modifications allowed us to discriminate between the resonant contribution and the NR background, as is explained below, and to extract the CARS signal even at zero probe delay. Furthermore, combining this method with the probe-pulse delay provided a means to suppress the interfering FWM and associated noise.

This combination of broadband preparation and frequency-resolved multichannel acquisition (*13–16*) with time-delayed narrowband probing yields a very sensitive and robust technique that allows us to identify bacterial endospores, such as anthrax, in real time. We refer to this technique as hybrid CARS for short, and we used it to study the Raman signature of endospores, which is dominated by the contribution from the dipicolinic acid (DPA) molecules.

In order to provide better insight into the advantages of the proposed scheme, we review here a few important aspects of the theory behind CARS, which has been extensively described elsewhere (*1–3, 8, 11, 14, 17*). In general, the third-order polarization induced by the pump, Stokes, and probe pulses can be split into resonant and NR contributions. If we assume that there are no one-photon resonances involved, the NR part can be attributed to the instantaneous electronic response and a sum over the contributions from far-detuned Raman transitions. The resonant part of the third-order polarization is attributed to the Raman transitions of interest. The NR component, $\chi_{NR}^{(3)}$, of the nonlinear susceptibility is usually frequency-insensitive within the considered spectral band and can be treated as a constant. The resonant component $\chi_{R}^{(3)}$, under the assumption of Lorentz-shape Raman lines, can be presented as $\chi_{R}^{(3)}(\omega_1 - \omega_2) = \sum_j A_j \Gamma_j / [\Omega_j - (\omega_1 - \omega_2) - i\Gamma_j]$, where ω_1 and ω_2 are the pump and Stokes laser

[1]Institute for Quantum Studies and Departments of Physics and Chemical Engineering, Texas A&M University, College Station, TX 77843, USA. [2]Applied Physics and Materials Science Group, Engineering Quadrangle, Princeton University, Princeton, NJ 08544, USA.

*To whom correspondence should be addressed. E-mail: dmip@neo.tamu.edu

Fig. 1. Level diagram and schematic of different scattering processes on simple molecules. In this example, CO is a target molecule and SO is a background molecule. (**A**) Incoherent Raman scattering (pulse 2) was derived from laser pulse 1 scattering off of the CO molecule. (**B**) CARS signal 4 was derived from probe pulse 3 scattering off of the CO molecular vibration, coherently prepared by pulses 1 and 2. (**C**) One of the possible channels for the NR background generation in SO. c, the ground state of the CO molecule; b, target vibrational state of the CO molecule; c′, the ground state of the background molecule; b′, an off-resonant vibrational state of the background molecule.

frequencies, j is the summation index over all covered Raman transitions, A_j is a constant related to the spontaneous Raman cross section of the jth Raman transition and molecular density, Γ_j denotes the jth Raman line half-width, Ω_j denotes the jth vibrational frequency (17), and $i = \sqrt{-1}$. The summation is held over all of the Raman transitions involved (17).

In the frequency domain (8), the third-order polarization $P^{(3)}(\omega)$ can be written as

$$P^{(3)}(\omega) = P_{NR}^{(3)}(\omega) + P_R^{(3)}(\omega) = \int_0^{+\infty} d\Omega \times [\chi_{NR}^{(3)} + \chi_R^{(3)}(\Omega)] E_3(\omega - \Omega) \times S_{12}(\Omega),$$ where $P_{NR}^{(3)}$ is the NR component of the polarization, $P_R^{(3)}$ is the resonant component, $E_3(\omega)$ is the spectral amplitude of the probe pulse, and $S_{12}(\Omega) \equiv \int_0^{+\infty} d\omega' \times E_1(\omega') E_2^*(\omega' - \Omega)$ is the convolution of the pump and Stokes field amplitudes, $E_1(\omega)$ and $E_2(\omega)$. The signal arising from the nonlinear response of the medium is proportional to $|P^{(3)}(\omega)|^2$, so the spectra generally have complex shapes caused by the interference between both resonant contributions from different vibrational modes and the NR background. A straightforward analogy with the spontaneous Raman spectra can be made only for well-separated lines with no NR background. Otherwise, the direct fit of the recorded CARS spectrum is required for Raman spectrum retrieval (18).

The convolution of the pump and Stokes spectra, $S_{12}(\Omega)$, enters the two parts of the third-order polarization on equal grounds. It defines a Raman frequency band covered by the preparation pulses and is maximized for transform-limited ones; i.e., pulses with the constant spectral phase. The difference between the two contributions comes from the susceptibility and can be enhanced by the use of a properly shaped probe. One way to proceed is to modify the spectral phase of the probe pulse, as it was demonstrated by Oron *et al.* (9). Another way, which seems to be more robust and straightforward, is to shape its spectral amplitude, $|E_3(\omega)|$, as we did in our experiment. If a narrowband probe is applied together with the broadband transform-limited preparation pulses, the NR contribution inherits a smooth, featureless profile of $S_{12}(\Omega)$ with some characteristic width $\Delta\omega_{12}$, whereas the resonant part generates a set of narrow peaks (one for each excited vibrational mode) whose width is determined either by the Raman linewidth or the probe spectral width, $\Delta\omega_3$, whichever is greater.

The amplitude ratio between the resonant signal and the NR background at a Raman-shifted frequency is also affected by the spectral width of the probe pulse. At the zero probe delay, the ratio is inversely proportional to the square of the probe spectral width when the last one is in between the Raman linewidth and the width of the pump-Stokes convolution profile (i.e., $\Gamma \ll \Delta\omega_3 \ll \Delta\omega_{12}$) and the probe-pulse energy is fixed. This ratio saturates at the limits

and leads to a superior but finite signal-to-background ratio for the optimum probe width on the order of the Raman linewidth.

The present measurement strategy combines the benefits of frequency-resolved CARS signal discrimination against the NR FWM (pointed out above) with the NR background suppression {as in time-resolved CARS [(19–22) and references therein]}. Indeed, when the probe delay is adjustable, further optimization is possible. In the plane of two parameters (the probe-pulse duration and its delay), the resonant response peaks for both parameters on the order of the inverse Raman linewidth. However, the NR FWM at the Raman-shifted frequency is maximized for zero probe delay, and the probe-pulse duration is matched to the time span of the pump-Stokes convolution profile.

As mentioned above, we can eliminate the NR background by just delaying the probe pulse, which gives a theoretically unlimited signal-to-background ratio. Unfortunately, this approach does not properly optimize the resonant contribution, and we can end up with no detectable signal at all. We suggest the simultaneous use of the two parameters (the probe-pulse duration and its delay) to achieve close-to-optimal resonant response with reasonable suppression of the NR background. The actual optimal values of the parameters depend on the Raman linewidth, the sensitivity of the setup used, and the relative strength of the resonant and NR susceptibilities.

Proper tailoring of the probe pulse can help to reduce the contribution of the NR background for probe delays comparable to the pulse length. For example, a rectangular-like spectrum gives a sinc-
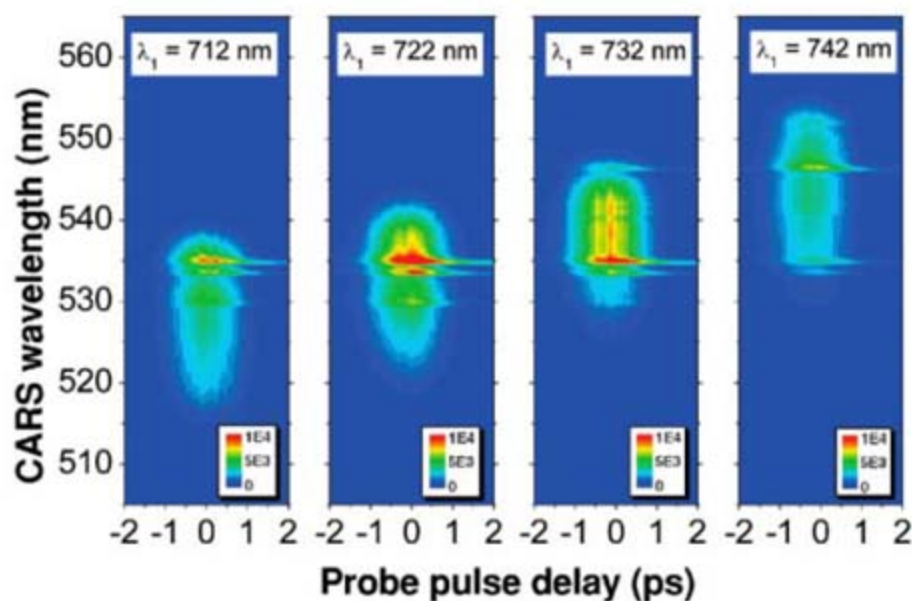
Fig. 2. CARS spectrograms recorded on NaDPA powder at different pump wavelengths. The CARS spectrum is shown as a function of the probe-pulse delay for the pump wavelength $\lambda_1 = 712$, 722, 732, and 742 nm, respectively (left to right). The other parameters are: pump, full width at half maximum (FWHM) ~12 nm, 2 μJ per pulse; Stokes, $\lambda_2 = 803$ nm, FWHM ~32 nm, 3.9 μJ per pulse; probe, $\lambda_3 = 577.9$ nm, FWHM ~0.7 nm, 0.5 μJ per pulse. The integration time was 1 s per probe-delay step. $1E4 = 1 \times 10^4$.

Fig. 3. Cross sections of the CARS spectrograms from Fig. 2 for two probe delays, (**A**) 0 ps and (**B**) 1.5 ps. The wavelengths within the observed range were transferred into the Raman shift, relative to the probe central frequency. The integration time was 1 s: 0.5 s for the signal and 0.5 s for the background acquired for the delayed Stokes pulse. The absolute frequencies of the Raman transitions in NaDPA, observed in the CARS experiment and spontaneous Raman measurements, are summarized in Table 1. CCD, charge-coupled device.
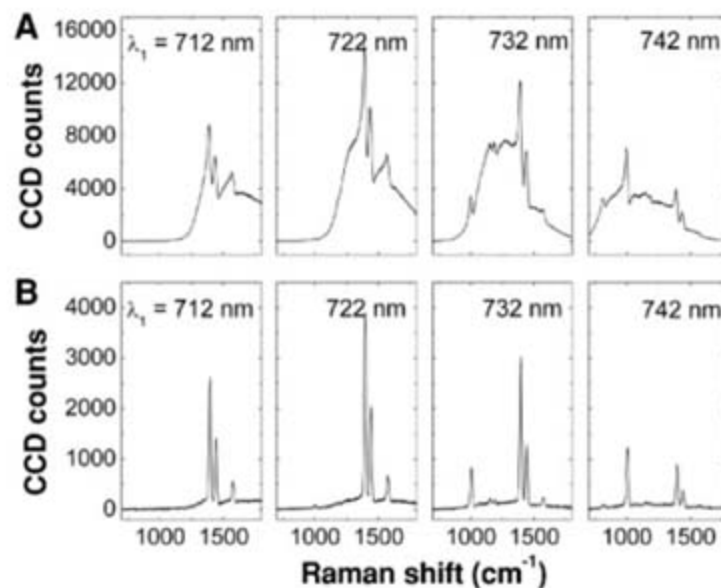
squared temporal profile of the probe-pulse intensity, which goes as $[\sin(\Delta\omega_3 t/2)/(\Delta\omega_3 t/2)]^2$ with the time t. Putting the preparation pulses in one of its nodes would result in the effective suppression of the NR background.

We compromise between the resolution, signal strength, and the extent of the NR background suppression. On a single-shot basis, the spectral resolution is usually determined by the probe bandwidth. However, this is not an intrinsic limit, and much better resolution can be achieved by recording the anti-Stokes spectrum while varying the probe-pulse delay, if the measurements are not overwhelmed by the fluctuations.

As mentioned above, we applied the developed technique in the context of the spore detection problem. A marker molecule for bacterial spores is calcium dipicolinate (CaDPA), which accounts for 10 to 17% of the bacterial spore dry weight (12). We focus here on NaDPA, which is easier to make and is a good surrogate for CaDPA. The spontaneous Raman spectrum of NaDPA exhibits a similar set of strong Raman lines as that of CaDPA. Both differ somewhat from the Raman spectrum of DPA itself (23). The important point is that although endospores are fairly complex in structure, their Raman spectra are dominated by several vibrational modes of CaDPA.

The details on our implementation of the hybrid CARS technique and the setup schematics can be found online (24). The CARS spectra of NaDPA powder as a function of the probe-pulse delay are shown in Fig. 2. The spectrograms were taken at different pump wavelengths to cover the whole spectral-fingerprint region of the molecule (800 to 1700 cm^{-1}). Streaklike horizontal lines are the signature of excited NaDPA Raman transi-

tions. The broadband pedestal is the NR background. As expected, the tuning of the pump wavelength spectrally shifts the NR FWM but leaves the position of the resonant lines untouched. Also, the resonant and NR contributions exhibit different dependencies on the probe delay. The magnitude of the NR background is determined by the overlap of the three laser pulses and follows the probe-pulse profile. However, a relatively long decay time of the Raman transitions under consideration favors their long-lasting presence and makes them stand out when the probe is delayed.

The cross sections of the spectrograms at two different probe delays are given in Fig. 3. The integration time for each of those is only 1 s. When the three pulses are overlapped (zero delay), the resonant contribution is severely distorted by the interference with the NR FWM. Delaying the probe by 1.5 ps—that is, putting the preparation pulses into the first node of the sinc-shaped probe pulse—improves the signal-to-background ratio by at least one order of magnitude. In this case, the NR background suppression is limited by multiple scattering that scrambles the timing between the preparation and probe photons.

The absolute frequencies of the observed Raman transitions calculated from the retrieved peak positions and the probe wavelength are summarized in Table 1. Comparison with the data from spontaneous Raman measurements shows a remarkably good match.

Extracted CARS contributions from our first measurements on *Bacillus subtilis* spores (a sur-

rogate for anthrax), in which we maximized the signal rather than the signal-to-background ratio, are summarized in Fig. 4. The Raman peaks are not normalized on the strength of the excitation and thus have an imprint of the pump-Stokes spectral convolution function, which sweeps through the Raman band from 800 to 1700 cm^{-1} while the pump wavelength is tuned. We assign the Raman transitions in the band (Table 1) and compare the retrieved line positions with the known positions from spontaneous Raman measurements (25, 26). Within the estimated experimental uncertainty of 15 cm^{-1}, the values are in good agreement. The data shown in Fig. 4 were acquired at zero probe delay over 2 min, although the Raman lines stand out from the background even after a few seconds of integration. Under similar experimental conditions, the signal arising from spontaneous Raman scattering is weaker by a few orders of magnitude and typically requires a longer integration time than the CARS signal.

To place the present work in context, our approach comes from the superposition of two well-known techniques developed over the past few decades and employed for combustion diagnostics [(1, 21, 22) and references therein] and chemically selective microscopic biological imaging (27, 28). Multifrequency acquisition has been implemented in so-called broadband or multiplex CARS (14–17), where together with the multichannel detection, a combination of narrowband pump-probe and broadband Stokes pulses is used to address a wide range of vibra-
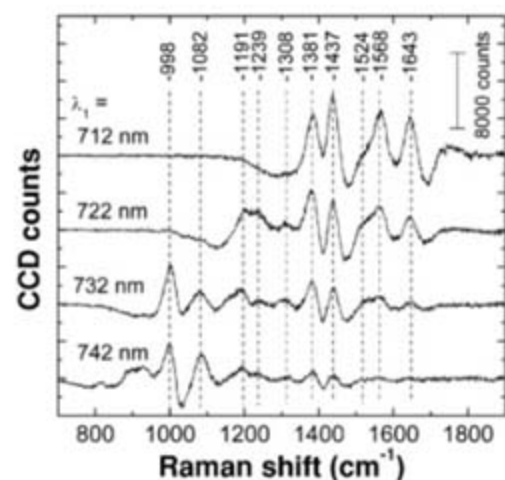
Fig. 4. CARS on *B. subtilis* spores at zero probe delay. The resonant contribution was retrieved by fitting the NR background with a smooth curve and subtracting it from the total acquired signal. The sample used was a pullet of spores fixed in a rotating sample holder. The pump wavelength, λ_1, was varied from 712 to 742 nm. Other parameters were the same as those for Fig. 2. The integration time was 2 min. The comparisons between the retrieved Raman frequencies and the available spontaneous Raman data are given in Table 1.

Table 1. The observed Raman peaks and their calculated absolute frequencies for NaDPA and *B. subtilis* spores. The third column lists the frequency values from spontaneous Raman measurements.

Peak (nm)	CARS Raman shift (cm^{-1})	Spontaneous Raman shift (cm^{-1})
	NaDPA powder	
529.7	1575	1572*
533.5	1440	1442*
534	1395	1395*
540.8	1187	1189*
541.8	1153	1152*, 1157*
543.9	1082	1087*
546.2	1004	1007*
551.9	815	817*, 827*
	B. subtilis spores	
527.8	1643	1655†, 1624‡
529.9	1568	1572†, 1581‡
531.2	1524	1539‡
533.6	1437	1445†, 1447‡
535.2	1381	1395†, 1396‡
537.3	1308	1280‡
539.3	1239	1245†
540.7	1191	1192‡
543.9	1082	
546.4	998	1001†, 1013† 981‡, 1018‡
551.8	819	822†

*See figure S2 in (24).   †From (25).   ‡From (26).

tional frequencies. In this degenerate scheme, the NR background has been addressed by means of the polarization-sensitive and interferometric techniques mentioned above.

A delayed probe has been used in time-resolved CARS (19–22). That technique uses ultrashort pulses for preparation and probing. Its ultimate source of species-specific information is multimode interference in the probe-delay signal profile, generally referred to as quantum beats (19). Time-resolved CARS eliminates the NR contribution by delaying the probe pulse, but the technique still remains vulnerable to fluctuations. It has been successfully applied to polycrystalline and opaque solids (29) to observe vibrational dephasing of single excited Raman transitions. However, the use of the multimode interference pattern for species recognition requires the ability to record high-quality quantum-beat profiles over a relatively large probe-delay span and therefore is challenging in the presence of scattering and fluctuations.

In our scheme, a generalized broadband or multiplex CARS technique is combined with background suppression by means of an optimal sequence of coherent excitation and time-delayed probe pulses. A schematic overview, pointing out the similarities and differences of the relevant established schemes and the one introduced here, is available online [figure S3 in (24)]. In short, we diverge from the conventional broadband CARS arrangement and deal with the probe and the two preparation pulses, pump and Stokes, separately. By adjusting the probe-pulse delay and its

spectral width, we suppress the NR background, as in time-resolved CARS, but keep the advantages of the frequency-resolved multiplex CARS spectroscopy. The experimental data demonstrate the efficacy of the ultrafast broadband excitation and time-variable narrowband probing, whereas the described implementation supports the versatility of the technique.

References and Notes

1. P. R. Régnier, J. P.-E. Taran, *Appl. Phys. Lett.* **23**, 240 (1973).
2. J. W. Nibler, G. V. Knighten, in *Raman Spectroscopy of Gases and Liquids*, A. Weber, Ed. (Springer-Verlag, New York, 1979), pp. 253–299.
3. T. G. Spiro, *Biological Applications of Raman Spectroscopy* (Wiley, New York, 1987).
4. W. M. Tolles, R. D. Turner, *Appl. Spectrosc.* **31**, 96 (1977).
5. J. L. Oudar, R. W. Smith, Y. R. Shen, *Appl. Phys. Lett.* **34**, 758 (1979).
6. Y. Yacoby, R. Fitzgibbon, B. Lax, *J. Appl. Phys.* **51**, 3072 (1980).
7. E. O. Potma, C. L. Evans, X. S. Xie, *Opt. Lett.* **31**, 241 (2006).
8. D. Oron, N. Dudovich, D. Yelin, Y. Silberberg, *Phys. Rev. A.* **65**, 043408 (2002).
9. D. Oron, N. Dudovich, D. Yelin, Y. Silberberg, *Phys. Rev. Lett.* **88**, 063004 (2002).
10. S. H. Lim, A. G. Caster, S. R. Leone, *Phys. Rev. A.* **72**, 041803 (2005).
11. T. W. Kee, H. Zhao, M. T. Cicerone, *Opt. Express* **14**, 3631 (2006).
12. M. O. Scully et al., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10994 (2002).
13. D. Klick, K. A. Marko, L. Rimai, *Appl. Opt.* **20**, 1178 (1981).
14. A. Voroshilov, C. Otto, J. Greve, *J. Chem. Phys.* **106**, 2589 (1997).
15. H. Kano, H. Hamaguchi, *Appl. Phys. Lett.* **85**, 4298 (2004).
16. G. I. Petrov, V. V. Yakovlev, *Opt. Express* **13**, 1299 (2005).
17. Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley, New York, 1984).
18. H. A. Rinia, M. Bonn, M. Muller, *J. Phys. Chem. B* **110**, 4472 (2006).
19. R. Leonhardt, W. Holzapfel, W. Zinth, W. Kaiser, *Chem. Phys. Lett.* **133**, 373 (1987).
20. A. Materny et al., *Appl. Phys. B* **71**, 299 (2000).
21. P. Beaud, H.-M. Frey, T. Lang, M. Motzkus, *Chem. Phys. Lett.* **344**, 407 (2001).
22. R. P. Lucht, S. Roy, T. R. Meyer, J. R. Gord, *Appl. Phys. Lett.* **89**, 251112 (2006).
23. P. Carmona, *Spectrochim. Acta A* **36**, 705 (1980).
24. The details of the experimental setup, a schematic overview of the established schemes (time-resolved CARS and multiplex/broadband CARS), and the technique introduced in this work are available as supporting material on *Science* Online.
25. A. P. Esposito et al., *Appl. Spectrosc.* **57**, 868 (2003).
26. W. H. Nelson, R. Dasari, M. Feld, J. F. Sperry, *Appl. Spectrosc.* **58**, 1408 (2004).
27. M. D. Duncan, J. Reintjes, T. J. Manuccia, *Opt. Lett.* **7**, 350 (1982).
28. A. Zumbusch, G. R. Holton, X. S. Xie, *Phys. Rev. Lett.* **82**, 4142 (1999).
29. X. Wen, S. Chen, D. D. Dlott, *J. Opt. Soc. Am. B* **8**, 813 (1991).
30. We thank J. Laane and K. McCann for their generous assistance with spontaneous Raman measurements and gratefully acknowledge support from the Office of Naval Research (award N00014-03-1-0385), the Defense Advanced Research Projects Agency, NSF (grant PHY-0354897), an award from the Research Corporation, and the Robert A. Welch Foundation (grants A-1261 and A-1547).

Designed Synthesis of 3D Covalent Organic Frameworks

Hani M. El-Kaderi,[1] Joseph R. Hunt,[1] José L. Mendoza-Cortés,[1] Adrien P. Côté,[1] Robert E. Taylor,[1] Michael O'Keeffe,[2] Omar M. Yaghi[1]*

Three-dimensional covalent organic frameworks (3D COFs) were synthesized by targeting two nets based on triangular and tetrahedral nodes: **ctn** and **bor**. The respective 3D COFs were synthesized as crystalline solids by condensation reactions of tetrahedral *tetra*(4-dihydroxyborylphenyl) methane or *tetra*(4-dihydroxyborylphenyl)silane and by co-condensation of triangular 2,3,6,7,10,11-hexahydroxytriphenylene. Because these materials are entirely constructed from strong covalent bonds (C-C, C-O, C-B, and B-O), they have high thermal stabilities (400° to 500°C), and they also have high surface areas (3472 and 4210 square meters per gram for COF-102 and COF-103, respectively) and extremely low densities (0.17 grams per cubic centimeter).

The chemistry of linking organic molecules together by means of covalent bonds to isolate crystals of discrete zero-dimensional (0D) molecules and 1D chains (polymers) is well established; however, it is undeveloped for 2D and 3D COFs (1). COF structures that contain light elements (B, C, N, and O) should be able to combine the thermodynamic strength of covalent bonds, as those found in diamonds and boron carbides, with the functionality of organic units. Progress in this area has been impeded by long-

standing practical and conceptual challenges. First, unlike 0D and 1D systems, the insolubility of 2D and 3D structures precludes the use of stepwise synthesis and makes their isolation in crystalline form very difficult. Second, the number of possible structures that may result from linking specific building-unit geometries into 2D or 3D extended structures is essentially infinite and complicates their synthesis by design.

We recently illustrated how the first challenge could be overcome by judiciously choosing

building blocks and using reversible condensation reactions to crystallize 2D COFs in which organic building blocks are linked entirely by strong covalent bonds (2). Here we report how the design principles of reticular chemistry overcome the second challenge (3): Two nets based on the linking of triangular and tetrahedral shapes were selected and targeted for the synthesis of four 3D COFs.

Self-condensation and co-condensation reactions of the rigid molecular building blocks, the tetrahedral *tetra*(4-dihydroxyborylphenyl)methane (TBPM), and its silane analog (TBPS), and triangular hexahydroxytriphenylene (HHTP) (Fig. 1, A to C) provided crystalline 3D COFs (termed COF-102, COF-103, COF-105, and COF-108). These COFs are the most porous among organic materials, and a member of this series (COF-108) has the lowest density reported of any crystalline material. Without our a priori

[1]Center for Reticular Chemistry at California NanoSystems Institute, Department of Chemistry and Biochemistry, University of California, Los Angeles, 607 Charles E. Young Drive East, Los Angeles, CA 90095, USA. [2]Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA.

*To whom correspondence should be addressed. E-mail: yaghi@chem.ucla.edu

knowledge of the expected underlying nets of these COFs, their synthesis by design and the solution of their structures from powder x-ray diffraction (PXRD) data would have been prohibitively difficult.

In planning the synthesis, we chose the tetrahedral building blocks (Fig. 1, A and B) and the triangular unit (Fig. 1C), because they are rigid and unlikely to deform during the assembly reaction. Dehydration reactions of these units produce triangular B_3O_3 rings and C_2O_2B rings (Fig. 1, D and E). Based on these building blocks, we envisioned two kinds of reactions in which

either of the tetrahedral blocks (Fig. 1, A and B) undergoes self-condensation or co-condensation with the triangular unit (Fig. 1C) to give COF structures based on nets with both tetrahedral and triangular nodes (Fig. 1, D and E).

In principle, there are an infinite number of possible nets that may result from linking tetrahedra with triangles. However, our analysis of previous assembly reactions suggests that the most symmetric nets are the most likely to result in an unbiased system and that those with just one kind of link will be preferred and are thus the best to target (3). In the present case of linking

tetrahedral and triangular building blocks, the only known nets meeting the above criteria are those with symbols **ctn** and **bor** (Fig. 1, F and G) (4). The nodes of the nets are thus replaced by the molecular building units with tetrahedral and triangular shapes (Fig. 1, H and I). The use of rigid, planar triangular units, such as B_3O_3 rings, requires that rotational freedom exist at the tetrahedral nodes for the 3D structures **ctn** and **bor** to form.

We then used Cerius² software to draw the "blueprints" for synthesis of COFs based on **ctn** and **bor** nets by fitting molecular building blocks (Fig. 1, A and B) on the tetrahedral nodes and by fitting the triangular unit and the B_3O_3 ring (Fig. 1, C and D) on the triangular nodes of these nets adhering to their respective cubic space group symmetries: $I\bar{4}3d$ (**ctn**) and $P\bar{4}3m$ (**bor**). Energy minimization by means of force-field calculations was performed to produce the models in which all bond lengths and angles were found to have chemically reasonable values (5).

Synthesis of the COFs was carried out by suspending either TBPM or TBPS in mesitylene/dioxane. The suspensions were placed in partially evacuated (150 mtorr) Pyrex tubes, which were sealed and heated (85°C) for 4 days to give white crystalline COF-102 and COF-103 in 63 and 73% yields, respectively. Similarly, co-condensation of TBPM or TBPS with HHTP (3:4 molar ratio) produced green crystalline solids of COF-105 (58% yield) and COF-108 (55% yield) (6). The colors of COF-105 and COF-108 likely arise from the possible inclusion of a small amount of highly colored oxidized HHTP in their pores. The use of dioxane and mesitylene in their respective ratios was necessary to control the solubility of the starting materials and to maximize crystallinity of the products.

To show that the products of synthesis are indeed covalently linked into the designed structures, we studied the materials by PXRD, spectroscopy, microscopy, elemental microanalysis, and gas adsorption (6). A comparison of PXRD patterns of modeled COFs to those observed for the products of synthesis (Fig. 2, A to D) reveals that they are indeed the expected COFs with **ctn** or **bor** type. The observed PXRD patterns display narrow line widths and low signal-to-noise ratios indicative of the high crystallinity of COFs. A marked degree of correspondence between peak positions and intensities is also observed, substantiating that the H, B, C, and O atomic composition and positions in the respective modeled unit cells are correct. The PXRD data of the COFs could also be indexed, yielding unit cell parameters nearly identical to those calculated from Cerius² (table S5).

To further verify the unit cell parameters, PXRD patterns were subjected to model-biased Le Bail full pattern decomposition to extract the structure factor amplitudes from the x-ray data. For this procedure to be successful and yield acceptable reliability factors, a close correspondence in peak position and intensity between the model and the experimental data is required. All
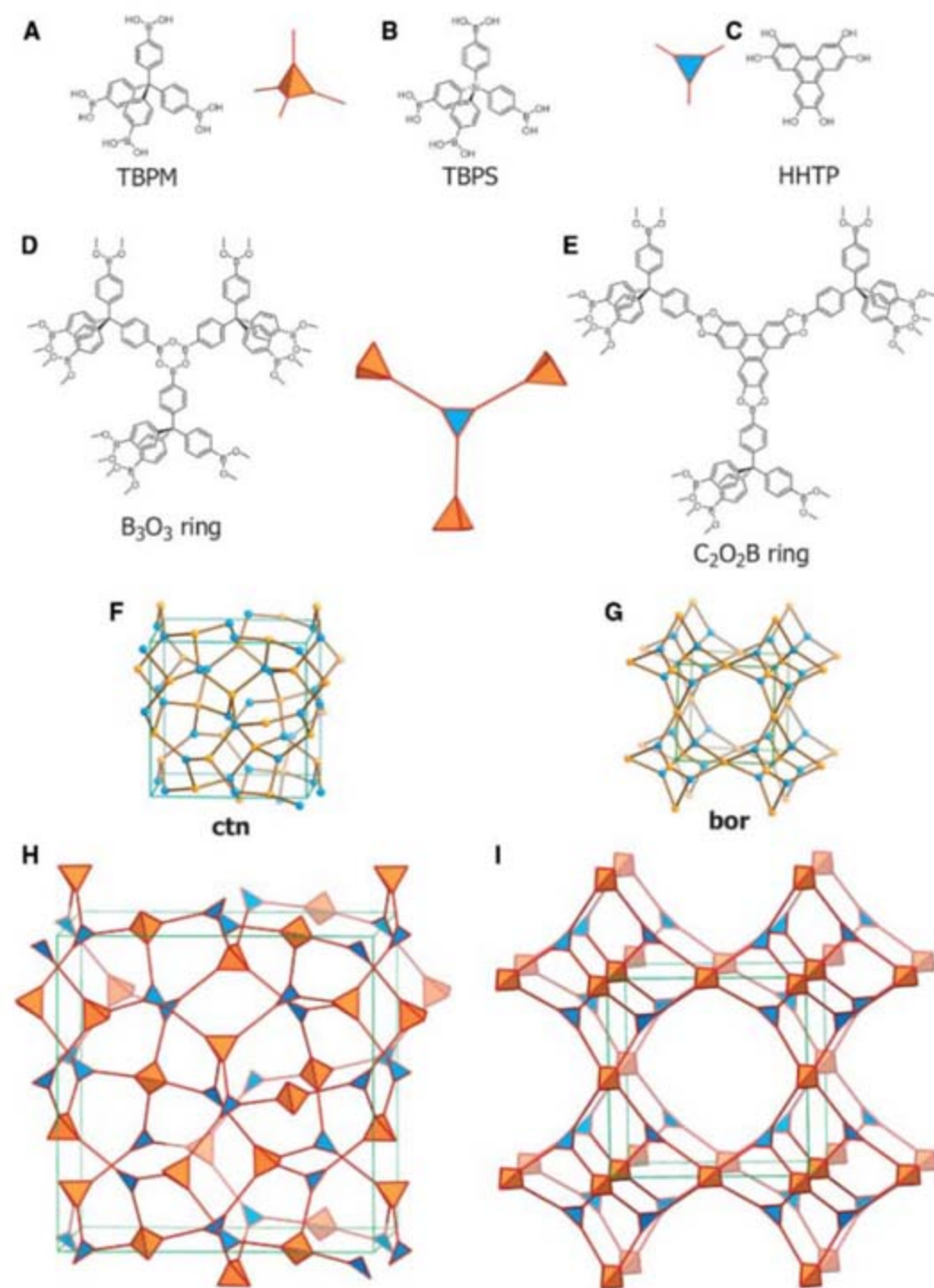
Fig. 1. Representative condensation routes to 3D COFs. Boronic acids are shown as tetrahedral building units [(**A**) and (**B**)], and a planar triangular unit (**C**) is also shown (polyhedron in orange and triangle in blue, respectively), including fragments revealing the B_3O_3 (**D**) and the C_2O_2B (**E**) ring connectivity in the expected linked products. These building units can be placed on the **ctn** (**F**) and **bor** (**G**) nets, as shown in the corresponding expanded nets (**H**) and (**I**), respectively.

peaks undergo some broadening because COF crystallites have micrometer dimensions (7). After accounting for line broadening in the initial stages of Le Bail extractions, fitting of the experimental profiles readily converged with refinement of the unit cell parameter, *a*. Refinements for all structures led, again, to values nearly identical to those calculated from Cerius[2] (table S5). Too few peaks were available to perform full Rietveld refinement of atomic positions and thermal parameters. Nonetheless, a near equivalence and low uncertainty [estimated SD (table S5)] between calculated and refined cell parameters, in addition to the facile and proper fit of the refined profiles [as indicated by statistically acceptable residual factors (table S6)], support the assertion that the COF structures are indeed those identified through modeling (Fig. 2; atomic coordinates in tables S1 to S4).

The covalent linking of building blocks through expected six-membered B_3O_3 boroxine
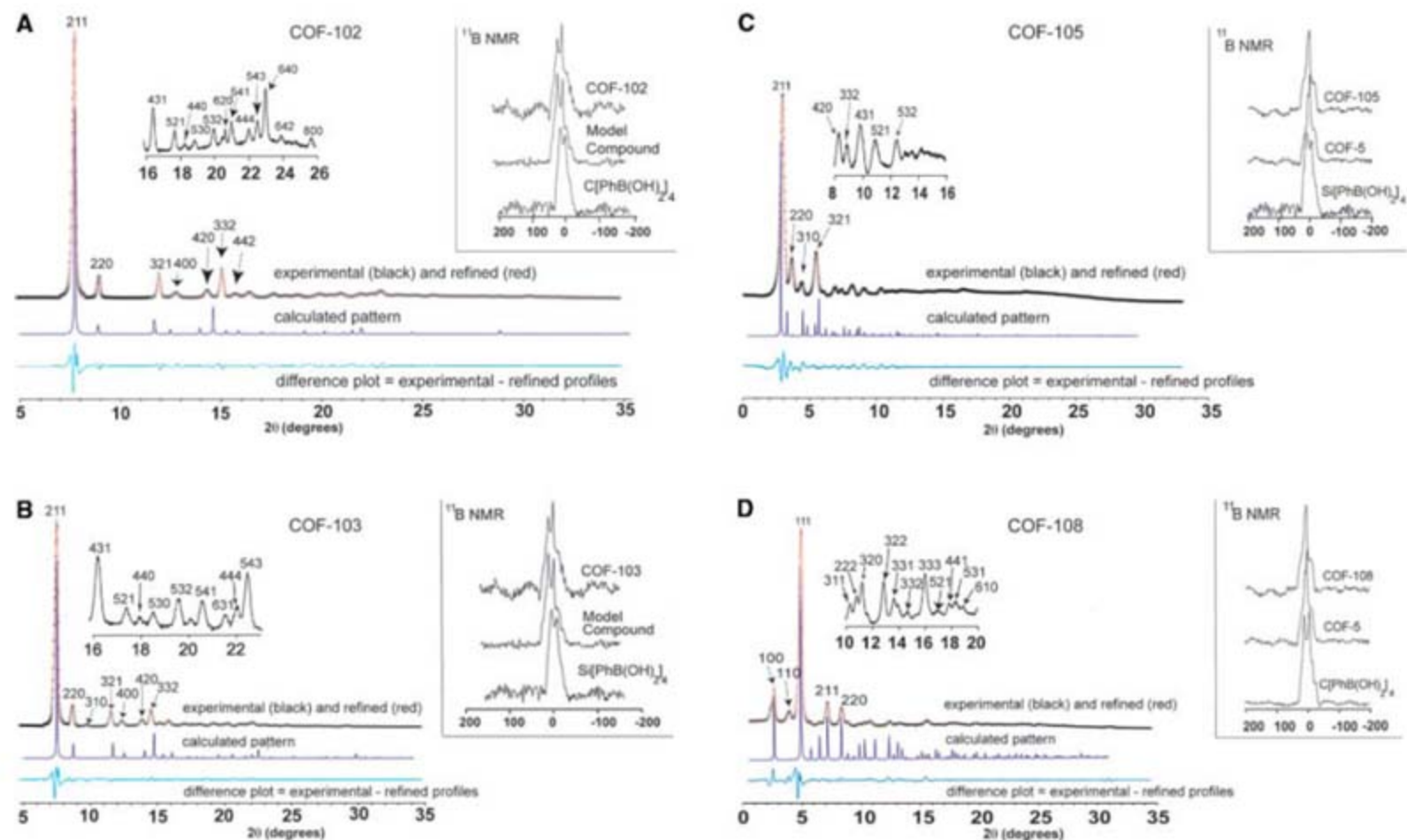
Fig. 2. Observed (experimental) and refined PXRD profiles for evacuated samples of COF-102 (**A**), COF-103 (**B**), COF-105 (**C**), and COF-108 (**D**), including patterns calculated with the use of Cerius[2], with observed profiles in black, refined profiles in red, calculated patterns in blue, and the difference plot (observed minus refined profiles) in turquoise. 2θ, Bragg angle in degrees. (Left insets) Expansion of observed PXRD profiles. (Right insets) ^{11}B MQ MAS NMR spectra (in parts per million) of COF (top trace), model compound (middle trace), and boronic acid (bottom trace) used to construct the corresponding COF.
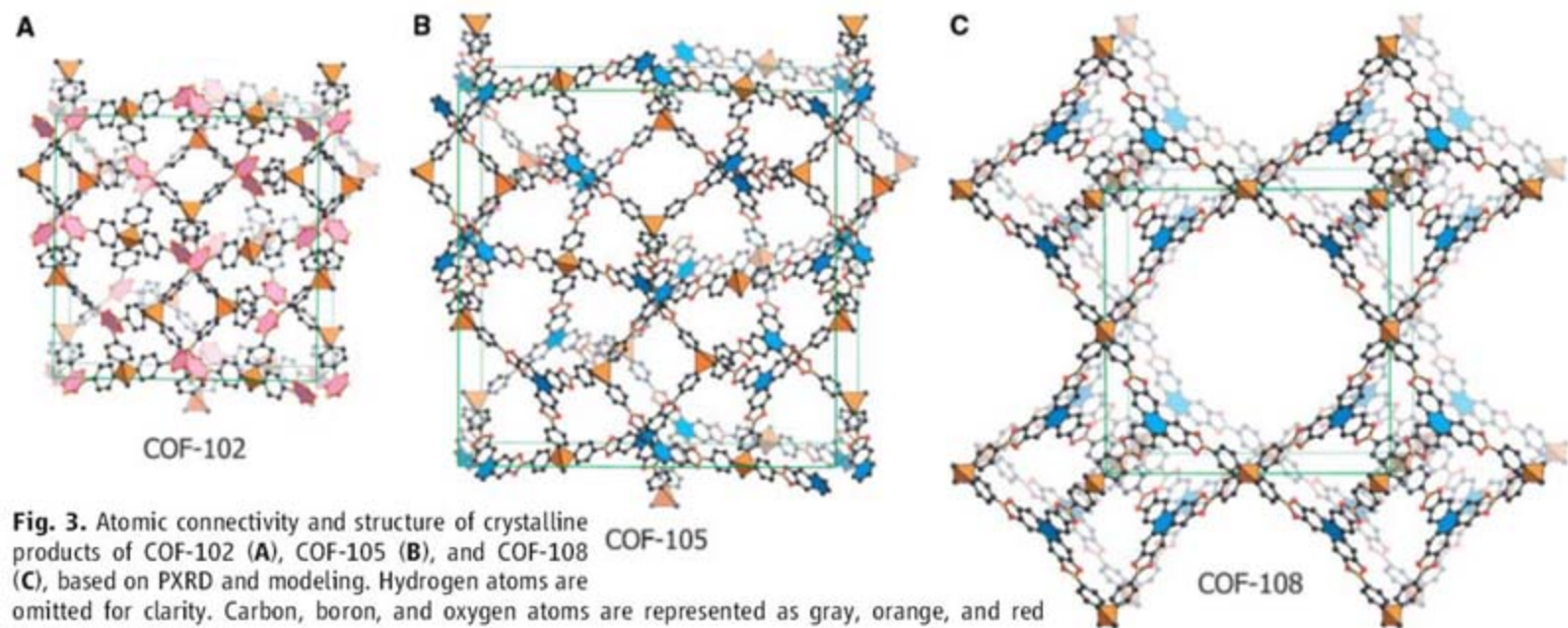
Fig. 3. Atomic connectivity and structure of crystalline products of COF-102 (**A**), COF-105 (**B**), and COF-108 (**C**), based on PXRD and modeling. Hydrogen atoms are omitted for clarity. Carbon, boron, and oxygen atoms are represented as gray, orange, and red spheres, respectively.

or five-membered C_2O_2B boronate ester rings in the COFs was assessed with Fourier transform infrared (FTIR) and multiple-quantum magic angle spinning nuclear magnetic resonance (MQ MAS NMR) spectroscopies. FTIR spectra of all COFs contain strongly attenuated bands arising from boronic acid hydroxyl groups indicative of successful condensation of the reactants (figs. S14 to S17). COFs prepared from self-condensation reactions all exhibit the diagnostic band at 710 cm^{-1} for the out-of-plane deformation mode of boroxine rings. Co-condensed COF-105 and COF-108 products have strong C-O stretching bands at 1245 cm^{-1} (COF-105) and 1253 cm^{-1} (COF-108), signals that are distinctive for boronate ester five-membered rings (6).

These FTIR spectroscopy data are fingerprints for the expected boron-containing rings; however, solid-state ^{11}B MQ MAS NMR spectroscopy is highly sensitive to the immediate bonding environment of boron. Any differences in B-C and B-O distances and/or angles will result in a notable change in the line shape and intensity of the spectra. The acquired ^{11}B MQ MAS NMR spectra for evacuated COFs were compared to those of molecular model compounds and starting materials (Fig. 2, A to D, right insets). The spectra of all of the COFs are coincident to those of the model compounds and are different from the starting materials. Thus, the boron-containing units in all the COFs have not only formed but are well-formed B_3O_3 and C_2O_2B rings. Additionally, data from ^{13}C and ^{29}Si MQ MAS NMR experiments show the presence of the expected number and environment of each type of respective nucleus, further substantiating the structural assignments (figs. S18 to S38).

In order to establish the phase purity and synthetic reproducibility of the COF materials, multiple samples were exhaustively imaged with scanning electron microscopy (SEM). The SEM images of COF-102 and COF-103 revealed agglomerated and nonagglomerated 1- to 2-μm-diameter spheres, respectively (figs. S39 and S40). This morphology is likely caused by a polar hydroxylated (-OH) surface that causes spherical crystal growth to minimize interfacial surface energy with the relatively nonpolar solvent media. The SEM images recorded for COF-105 and COF-108 revealed 5-μm platelets and 3- to 4-μm irregular spheres, respectively (figs. S41 and S42). For each of the COFs, only one morphology was observed, negating the presence of impurity phases. Furthermore, C and H elemental microanalysis confirmed that the composition of each COF corresponded to formulations predicted from modeling (6).

The derived structures for COF-102, COF-105, and COF-108 are shown in Fig. 3 (COF-103 has a tetrahedral Si replacing C, and its structure is virtually identical to that of COF-102). COF-102 (Fig. 3A), COF-103, and COF-105 (Fig. 3B) are based on **ctn**, and COF-108 (Fig. 3C) is based on **bor**. It is hard to assess why one of the two structure types would be preferred over the other. However, it is a notable confirmation of our original thesis that we find one or the other of the two structures. The only notable differences between the two structure types are that **bor** is about 15% less dense than **ctn** (compare the densities of COF-105 and COF-108) and has larger pores. The three-coordinated vertices in both structures are constrained to be planar with threefold symmetry, but the point symmetry at the tetrahedral site in **ctn** is only a subgroup ($\overline{4} = S_4$) of that at the tetrahedral site in **bor** ($\overline{4}m2 = D_{2d}$); this difference gives **ctn** less constraints and potentially makes it a more strain-free structure than **bor**.

It is also of interest to consider the pore sizes. In the COFs with the **ctn** structure, the center of the largest cavity in COF-102, COF-103, and COF-105 is 5.66, 5.98, and 10.37 Å, respectively, from the nearest atoms (H). If we allow for a van der Waals radius of 1.2 Å for H, spheres of diameter 8.9, 9.6, and 18.3 Å, respectively, are available in these three COFs. However, the pores in these materials are far from spherical, and we expect the effective pore size to be somewhat larger. COF-108 has two cavities, and the atoms closest to the center are C atoms at 9.34 and 15.46 Å. If we allow for a van der Waals radius of 1.7 Å for C, these cavities can accommodate spheres of 15.2 and 29.6 Å, respectively. It may be seen that the larger pores are well above the lower limit (20 Å) for the material to be described as mesoporous, and COF-108 is a rare example of a fully crystalline mesoporous material.

An important feature of 3D COFs is the full accessibility from within the pores to all the edges and faces of the molecular units used to construct the framework. A previous study found that maximizing the number of edges arising from aromatic rings in a porous material increases the number of adsorption sites and surface area (8). Porous zeolites, carbons, and metal-organic frameworks (MOFs) all contain latent edges in their structures; however, the structures of COFs contain no latent edges, and the entire framework is a surface replete with binding sites for gas adsorption. The structures also have extraordinarily low densities: COF-102 (0.41 g cm^{-3}), COF-103 (0.38 g cm^{-3}), COF-105 (0.18 g cm^{-3}), and COF-108 (0.17 g cm^{-3}). The last two values are markedly lower than those of highly porous MOFs such as MOF-5 (0.59 g cm^{-3}) (9) and MOF-177 (0.42 g cm^{-3}) (8) and are the lowest density crystals known (10) [compare these values with the density of diamonds (3.50 g cm^{-3})].

The low densities coupled with the maximized fraction of surface sites in 3D COFs naturally impart their exceptional porosities, which were shown in gas adsorption studies on evacuated samples of COF-102 and COF-103. Samples of "as synthesized" COF-102 and COF-103 were immersed in anhydrous tetrahydrofuran to remove solvent and starting materials
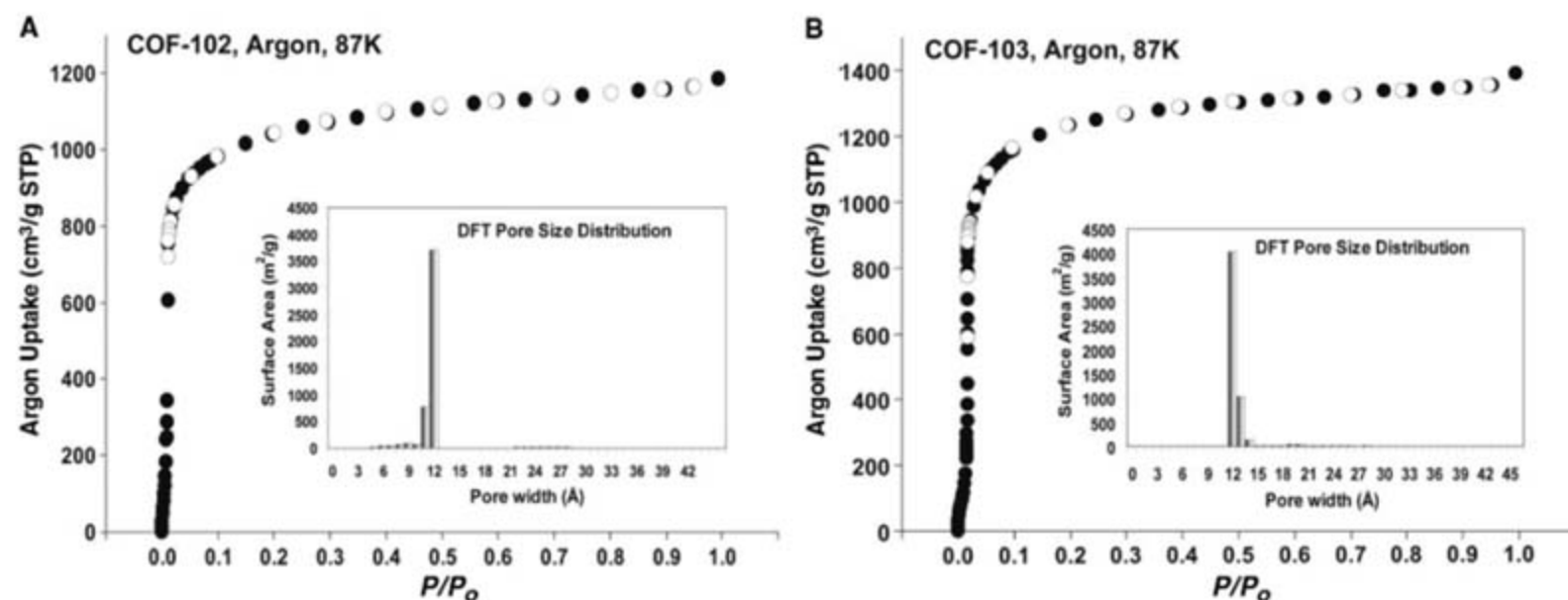
Fig. 4. Argon gas adsorption isotherms for COF-102 (**A**) and COF-103 (**B**) measured at 87 K and pore size histograms (insets) calculated after fitting DFT models to gas adsorption data. STP, standard temperature and pressure.

included in the pores during synthesis and were then placed under dynamic vacuum conditions (10^{-5} torr) for 12 hours at 60°C to completely evacuate the pores (6). Thermogravimetric analysis confirmed that all of the guests were removed from the pores and revealed the thermal stability of all COFs beyond 450°C (figs. S43 and S46). Argon isotherms for COF-102 and COF-103 were recorded at 87 K from 0 to 760 torr (Fig. 4, A and B). COF-102 and COF-103 exhibit a classic type I isotherm characterized by a sharp uptake at the low-pressure region between $P/P_o = 1 \times 10^{-5}$ to 1×10^{-2}, where P is gas pressure and P_o is saturation pressure. The apparent surface areas calculated from the Brunauer-Emmett-Teller (BET) model were 3472 and 4210 $m^2\ g^{-1}$ for COF-102 and COF-103, respectively. The pore volume determined from the Dubinin-Radushkevich equation afforded values of 1.35 $cm^3\ g^{-1}$ (COF-102) and 1.66 $cm^3\ g^{-1}$ (COF-103). The BET surface areas of COFs exceed porous carbons (2400 $m^2\ g^{-1}$) (11), silicates (1300 $m^2\ g^{-1}$) (12), recently reported 2D COFs (1590 $m^2\ g^{-1}$) (2), polymers of intrinsic microporosity (1064 $m^2\ g^{-1}$) (13), and polymer resins (2090 $m^2\ g^{-1}$) (14) and are comparable to some of the highest surface areas of MOFs [MOF-177 (4500 $m^2\ g^{-1}$) (8) and MIL-101 (4100 $m^2\ g^{-1}$) (15) (MIL, Matérial Institut Lavoisier)]. Calculation of pore size obtained from appropriately fitting density functional theory (DFT) models to the isotherms (figs. S48

and S52) yielded pore size distributions of COF-102 (11.5 Å) (Fig. 4A, inset) and COF-103 (12.5 Å) (Fig. 4B, inset) (16). Narrow distributions are obtained and are centered at values close to the pore diameters obtained from the crystal structures.

At the outset of this study, crystallization of 3D COFs (such as cross-linked polymers) was believed to be difficult, if not impossible, to achieve for both thermodynamic and kinetic reasons. This report demonstrates that this challenge can be met by striking a balance between these two competing factors and that the principles of reticular chemistry provide the basis for design and structure solution of the resulting materials.

References and Notes

1. N dimensional refers to materials indefinitely extended in N independent directions.
2. A. P. Côté et al., Science 310, 1166 (2005).
3. Reticular chemistry is concerned with the linking of molecular building blocks into predetermined structures by strong bonds (17).
4. O. Delgado-Friedrichs, M. O'Keeffe, O. M. Yaghi, Acta Crystallogr. A 62, 350 (2006).
5. Cerius² Modeling Environment, version 4.2, Molecular Simulations Incorporated, San Diego, CA (1999).
6. A full description of the synthesis, characterization, and gas adsorption measurements can be found in the supporting material on Science Online.
7. B. D. Cullity, Elements of X-ray Diffraction (Addison-Wesley, Don Mills, Ontario, ed. 2, 1978).
8. H. K. Chae et al., Nature 427, 523 (2004).
9. H. Li, M. Eddaoudi, M. O'Keeffe, O. M. Yaghi, Nature 402, 276 (1999).
10. The Cambridge Structural Database reports two structures with densities less than 0.18 g cm^{-3}. However, these reports arise from data deposited for incomplete structures where all atoms have not been included in the calculation of density.
11. N. Texier-Mandoki et al., Carbon 42, 2744 (2004).
12. M. Thommes, in Nanoporous Materials Science and Engineering, G. Q. Lu, X. S. Zhao, Eds. (Imperial College Press, London, 2004).
13. B. S. Ghanem et al., Chem. Commun. 2007, 67 (2007).
14. J.-H. Ahn et al., Macromolecules 39, 627 (2006).
15. G. Férey et al., Science 309, 2040 (2005).
16. K. Schumacher, P. I. Ravikovitch, A. D. Chesne, A. V. Neimark, K. K. Unger, Langmuir 16, 4648 (2000).
17. O. M. Yaghi et al., Nature 423, 705 (2003).
18. The work was supported by Badische Anilin und Soda Fabrik (BASF) Ludwigshafen for synthesis, U.S. Department of Energy (DEFG0206ER15813) for adsorption, and NSF (DMR 0242630) for simulated structures. We thank H. Furukawa (Yaghi group) for invaluable assistance with porosity measurements. Crystallographic information files for COF-102, COF-103, COF-105, and COF-108 have been deposited into the Cambridge Crystallographic Data Centre (CCDC) under deposition numbers CCDC 637175 to 637178.

Direct Detection of the Asteroidal YORP Effect

Stephen C. Lowry,[1]* Alan Fitzsimmons,[1] Petr Pravec,[2] David Vokrouhlický,[3] Hermann Boehnhardt,[4] Patrick A. Taylor,[5] Jean-Luc Margot,[5] Adrian Galád,[6] Mike Irwin,[7] Jonathan Irwin,[7] Peter Kusnirák[2]

The Yarkovsky-O'Keefe-Radzievskii-Paddack (YORP) effect is believed to alter the spin states of small bodies in the solar system. However, evidence for the effect has so far been indirect. We report precise optical photometric observations of a small near-Earth asteroid, (54509) 2000 PH5, acquired over 4 years. We found that the asteroid has been continuously increasing its rotation rate ω over this period by $d\omega/dt = 2.0\ (\pm 0.2) \times 10^{-4}$ degrees per day squared. We simulated the asteroid's close Earth approaches from 2001 to 2005, showing that gravitational torques cannot explain the observed spin rate increase. Dynamical simulations suggest that 2000 PH5 may reach a rotation period of ~20 seconds toward the end of its expected lifetime.

The Yarkovsky-O'Keefe-Radzievskii-Paddack (YORP) effect is a torque that can modify the rotation rates and obliquities of small bodies in the solar system; its causes are incident solar radiation pressure and the recoil effect from anistropic emission of thermal photons (1). Several effects indicate that such a torque, as yet undetected, acts upon the surfaces of asteroids and meteoroids, and the YORP effect is the only realistic mechanism in these cases. A clear example of a YORP-evolved system can be seen within the Koronis asteroid family, formed by the catastrophic collisional disruption of a large parent

body at least 2 billion years ago. This event would presumably have resulted in random spin states for the daughter asteroids, but surprisingly, the larger members are divided into two distinct alignments (2). Theoretical modeling invoked YORP as a potential mechanism to explain the spin-vector alignments by eventually altering the spin rates and obliquities, with some asteroids becoming trapped in spin-orbit resonances (3).

Another likely manifestation of YORP torques is the evolution of orbital semimajor axes of small members of asteroid families toward extreme values, which is important for determining

their age (4). The evolving obliquity will have a direct bearing on the evolution of semimajor axes via Yarkovsky drift, a companion effect to YORP that is caused by a net force arising from emission of thermal radiation along the body's orbit and was recently detected for the first time (5). YORP is also an important component in models of the delivery of near-Earth asteroids from the main asteroid belt, as it leads to a Yarkovsky drift, forcing bodies into powerful perturbing orbital resonances (6).

Furthermore, there exists a distinct population of small, very-fast-spinning asteroids known as the monolithic fast rotators (MFRs) (7). If the orbits of such bodies are stable over million-year time scales, allowing YORP to significantly change the rotation rate, then the effect is highly

[1]School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK. [2]Astronomical Institute, Academy of Sciences of the Czech Republic, Fricova 1, CZ-25165 Ondrejov, Czech Republic. [3]Institute of Astronomy, Charles University, V Holesovickách 2, 18000 Prague 8, Czech Republic. [4]Max-Planck-Institut für Sonnensystemforschung, Max-Planck-Strasse 2, 37191 Katlenburg-Lindau, Germany. [5]Department of Astronomy, Cornell University, Ithaca, NY 14853, USA. [6]Department of Astronomy, Physics of the Earth, and Meteorology, Faculty of Mathematics, Physics and Informatics, Comenius University, 842 48 Bratislava, Slovakia. [7]Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK.

*To whom correspondence should be addressed. E-mail: s.c.lowry@qub.ac.uk

likely to place these asteroids into this category. YORP could also cause small asteroids to spin up so fast that they are forced to morph into new shapes or even shed mass; perhaps such a progression would supersede tidal disruption and collisions as the main formation mechanism for binary asteroids in the planet-crossing population (8, 9). Conversely, YORP could act against the rotation direction to reduce the spin rate, leading to the very long rotation periods (>40 days) seen for some asteroids and possibly causing some of them to tumble. Despite its importance, there exists only indirect evidence for the presence of YORP on solar system objects.

Although the obliquity effect is too small to be detected with ground-based or even space-based instruments, changes in the rotation rate may be detected if very precise observations, such as high-quality asteroid light curves, are obtained over a long enough time span (10, 11). The small near-Earth asteroid (54509) 2000 PH5 (hereafter PH5) is one of the few known Earth co-orbitals in a near 1:1 mean-motion resonance with Earth. The characteristics of its horseshoe orbit result in periodic close Earth approaches, roughly every July and August, making it a good target for regular monitoring over yearly time scales. Here, we present results from long-term optical photometric monitoring of PH5 that reveal the YORP effect in action via an observed decrease in its sidereal rotation period, which is shown in our companion paper (12) to be in reasonable agreement with theoretical predictions for the effect. In (12), the optical data are analyzed alongside radar data to reconstruct a detailed shape model from which the expected YORP torques can be reliably assessed.

With a small radar-derived mean radius of ~57 m, any size-related YORP effect on PH5 is likely to be relatively high. Its fast spin period of 12.17 min makes it a very practical target for observations from Earth-based telescopes, because full and consecutive nights of observation are not required for accurate period determination. In 2 hours we could continuously sample the light curve over ~10 full revolutions, more than adequate for our purposes. This small size and fast rotation puts PH5 into the MFR asteroid group, so YORP may have been significantly affecting its spin rate for some time, and it is

reasonable to assume that it is still susceptible to YORP torques.

Our optical monitoring campaign began in 2001 and spanned ~4 years, with many observatories from various countries contributing telescope time. Because of the close approaches of PH5, the apparent brightness of the asteroid as seen from Earth changed considerably over time scales of days to weeks. Therefore, as the integration time was restricted to 30 s to avoid rotational smearing of the light curve, a range of optical telescope apertures (0.65 to 8.2 m) was required to reach the desired signal-to-noise ratio in this time. The asteroid's rotational signature

Fig. 1. Light curve data from the 2002–2003 data group, folded to the best-fit sidereal period of 0.20290020 hours. The fortuitous near-constant viewing geometry throughout the observations ensured that the shape and amplitude variations in the light curve were small from one night to the next.
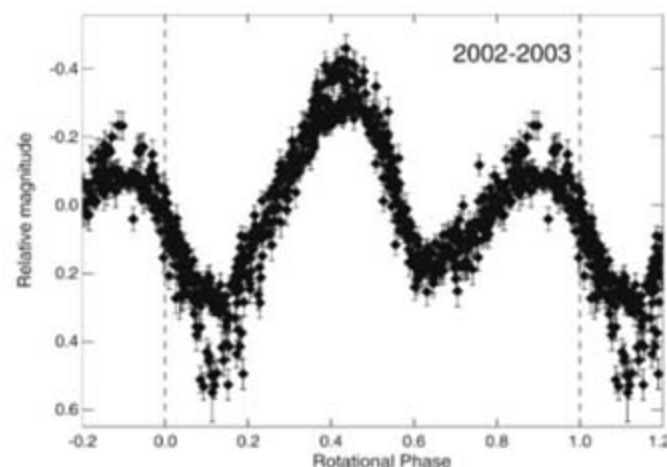
Fig. 2. Observed YORP effect on asteroid (54509) 2000 PH5. The sidereal rotation period P is steadily decreasing. For a solid YORP detection it is very important to confirm the existence of a trend, hence the need for repeated measurements over subsequent years. A linear trend is seen in $\Delta P/P_0 = (P - P_0)/P_0$ versus time, where $P_0 = 730.440756$ s (i.e., the nominal value from the 2002–2003 data group). Solid symbols represent the observed period change (the formal 1σ uncertainty on the period is smaller than the plotted data points). The solid curve is a theoretical prediction for YORP strength computed from different shape variants of the asteroid and recalibrated by a factor of 0.16 to 0.38, as discussed in (12). Open circles (with 1σ uncertainties) indicate the expected change of P during annual close encounters with Earth (arbitrarily offset for clarity), which is negligible relative to the observed period change.

Table 1. Evolution of the sidereal rotation period. There is partial overlap between successive data groups; however, the PAB corrections are performed on each biannual data group independently (13).

Date range of data group	Sidereal period (hours)*
28 Jul 2001–27 Jul 2002	0.20290046
27 Jul 2002–27 Aug 2003	0.20290020
29 Jul 2003–11 Sep 2004	0.20289985
12 Aug 2004–31 Aug 2005	0.20289941

*1σ uncertainties are 10^{-8} hours.

Fig. 3. Future orbital and spin evolution of asteroid (54509) 2000 PH5 from numerical integrations. Left: Number of surviving particles (right ordinate, curve 1) and median inclination I (left ordinate, curve 2) versus time. Right: Distribution of rotation period P values for the surviving particles after 1 My (93% of the population surviving), 5.5 My (75% surviving), 14.8 My (50% surviving), and 35 My (25% surviving). The spin-axis obliquity rarely drops below ~125°, and so YORP continues to increase the particle's rotation rate.

was extracted from each individual imaging data set (13), and the resulting light curves were grouped according to date; all the 2001 and 2002 light curves were taken together, then all the 2002–2003, 2003–2004, and 2004–2005 data, so that each data set had a 1-year time base. To find changes in the sidereal period, we shifted the observation times of the data points within each combined set by the established phase angle bisector (PAB) approximation (13, 14), which requires knowledge of the spin-axis orientation. Using a combination of radar data and our optical light curves, Taylor et al. (12) report that the pole position resides within 10° of (180°, −85°) in ecliptic coordinates (J2000).

Fourier analysis of the light curve modulation was then performed separately for each time-corrected data group (13) to determine the yearly averaged sidereal rotation periods (Table 1). The light curve amplitude changed little within any data set because of the nearly constant aspect angle of the asteroid throughout the observations. As a result of the asteroid's shape, there was also a pronounced asymmetry in the light curve that ultimately allowed unambiguous phasing of the data (Fig. 1). Sidereal rotation periods P were determined from 2001 to 2005 and were seen to decrease at a linear rate, with a fractional change of -1.7×10^{-6} ($\pm9\%$) per year; that is, the asteroid has been increasing its rotation rate ω over these 4 years by $d\omega/dt = 2.0$ (±0.2) $\times 10^{-4}$ degrees day^{-2} (Fig. 2). This result was confirmed from analysis of the combined light curve and radar data in (12). Detailed dynamical simulations that used the shape model in (12) were performed that reproduced the close Earth approaches from 2001 to 2005, from which we can rule out Earth-tug effects as a potential mechanism for the observed sidereal-period decrease (Fig. 2) (13). Moreover, there is no reason for Earth tugs to be coherent, so uncorrelated positive and negative shifts in spin rate are possible in subsequent years. The reasonable agreement between observations and YORP simulations (12) and the fact that planetary tugs cannot account for the observed effect leaves YORP as the only viable cause.

The fast rotation of PH5 could imply that this asteroid underwent significant YORP evolution in the past. Indeed, its obliquity near 180° supports this idea because it is near one of the asymptotic YORP regimes (11, 15). Our result suggests that it would take ~550,000 years for YORP to double the rotation rate of PH5 in the future. From this value we may expect that YORP will cause structural changes, mass shedding, or even fission of this object at some point in the future, depending on its internal strength. To investigate this possibility, we ran a simulation that numerically propagated the orbit of PH5 and 999 close clones (13). We found a median dynamical lifetime before particle removal from the simulation, by solar or planetary impacts, of ~15 million years (My), a surprisingly long time scale (Fig. 3). The longest-lived clones of PH5 (about 6%) survived 100 My of orbital evolution.

In a second step, we numerically integrated the secular evolution of the spin state for each of these 1000 particles along their precise orbits (13), with the YORP strength set to our observed value. At 35 My, when 25% of the original clone population remained, the median rotation period was 19 s with a lowest extreme of 5 s (Fig. 3).

Our observational calibration of the YORP effect, in conjunction with orbital and spin integrations, demonstrates that asteroids like PH5 can attain extremely fast rotation rates. Our work also implies the possible existence of a population of 100-m asteroids with rotation periods of ~20 s, significantly faster than the most rapidly rotating asteroid of this size, 2000 WH10 with $P \sim 80$ s (16). Light curve observations to date are biased against the detection of such short periods, and hence the number of such bodies is unconstrained at present. If no such objects are found, then the most likely explanation is eventual significant mass shedding or rotational fission before they reach this value of P.

References and Notes

1. D. P. Rubincam, *Icarus* **148**, 2 (2000).
2. S. M. Slivan, *Nature* **419**, 49 (2002).
3. D. Vokrouhlický, D. Nesvorný, W. F. Bottke, *Nature* **425**, 147 (2003).
4. D. Vokrouhlický, M. Broz, W. F. Bottke, D. Nesvorný, A. Morbidelli, *Icarus* **182**, 118 (2006).
5. S. R. Chesley et al., *Science* **302**, 1739 (2003).
6. A. Morbidelli, D. Vokrouhlický, *Icarus* **163**, 120 (2003).
7. P. Pravec, A. W. Harris, T. Michalowski, in *Asteroids III*, W. F. Bottke, A. Cellino, P. Paolicchi, R. Binzel, Eds. (Univ. of Arizona Press, Tucson, AZ, 2002), pp. 113–122.
8. W. F. Bottke, D. Vokrouhlický, D. P. Rubincam, M. Broz, in *Asteroids III*, W. F. Bottke, A. Cellino, P. Paolicchi, R. Binzel, Eds. (Univ. of Arizona Press, Tucson, AZ, 2002), pp. 395–408.
9. W. F. Bottke, D. Vokrouhlický, D. P. Rubincam, D. Nesvorný, *Annu. Rev. Earth Planet. Sci.* **34**, 157 (2006).
10. D. Vokrouhlický, D. Capek, M. Kaasalainen, S. J. Ostro, *Astron. Astrophys.* **414**, L21 (2004).
11. D. Capek, D. Vokrouhlický, *Icarus* **172**, 526 (2004).
12. P. A. Taylor et al., *Science* **316**, 274 (2007); published online 8 March 2007 (10.1126/science.1139038).
13. See supporting material on *Science* Online.
14. R. C. Taylor, in *Asteroids*, T. Gehrels, Ed. (Univ. of Arizona Press, Tucson, AZ, 1979), pp. 480–493.
15. D. Vokrouhlický, D. Capek, *Icarus* **159**, 449 (2002).
16. R. J. Whiteley, C. W. Hergenrother, D. J. Tholen, *Proc. ACM 2002*, 473 (2002).
17. We thank all the staff at the observatories involved in this study for their support. This work was based on observations at the following observatories: ESO, Chile (PIDs 271.C-5023 and 073.C-0137); Ondrejov Observatory, Czech Republic (grant A3003204); Centro Astronómico Hispano Alemán (Calar Alto, Spain); Liverpool Telescope, La Palma (Canary Islands, Spain); Isaac Newton Group, La Palma; and Faulkes Telescope North (Maui, Hawaii). We also thank the referees for their helpful reviews. Supported by the Leverhulme Trust (S.C.L.), the UK Particle Physics and Astronomy Research Council (A.F.), the Grant Agency of the Czech Republic (P.P. and D.V.), NASA grant NNG04GN31G (P.A.T. and J.-L.M.), and Slovak Grant Agency for Science VEGA grant 1/3074/06 (A.G.). This work made use of the NASA/JPL HORIZONS ephemeris-generating program.

Spin Rate of Asteroid (54509) 2000 PH5 Increasing Due to the YORP Effect

Patrick A. Taylor,[1]* Jean-Luc Margot,[1]* David Vokrouhlický,[2] Daniel J. Scheeres,[3] Petr Pravec,[4] Stephen C. Lowry,[5] Alan Fitzsimmons,[5] Michael C. Nolan,[6] Steven J. Ostro,[7] Lance A. M. Benner,[7] Jon D. Giorgini,[7] Christopher Magri[8]

Radar and optical observations reveal that the continuous increase in the spin rate of near-Earth asteroid (54509) 2000 PH5 can be attributed to the Yarkovsky-O'Keefe-Radzievskii-Paddack (YORP) effect, a torque due to sunlight. The change in spin rate is in reasonable agreement with theoretical predictions for the YORP acceleration of a body with the radar-determined size, shape, and spin state of 2000 PH5. The detection of asteroid spin-up supports the YORP effect as an explanation for the anomalous distribution of spin rates for asteroids under 10 kilometers in diameter and as a binary formation mechanism.

Theory predicts an evolution of the spin state of a small solar system body as a result of the absorption and asymmetric re-emission of sunlight, the so-called YORP effect. The resultant radiation torques are thought to realign the spin vector while changing the spin rate of the object (1). Lowry et al. (2) report an increase in the spin rate of 2000 PH5. Here we present, with radar and optical observations and modeling of 2000 PH5, the best

evidence to date that the YORP effect is responsible for changing the spin rate of an asteroid.

For objects with finite thermal conductivity, especially fast-rotating asteroids <1 km in diameter that lack regolith, YORP torques tend to force the spin vector to 0° or 180° obliquity (parallel or antiparallel to the orbit normal) and cause the spin rate to increase or decrease with equal probability (3) on time scales proportional to the square of the diameter D (1). Therefore, YORP may explain

the observed excess of slow and rapid rotators among asteroids <10 km in diameter (*4*). Continuous spin-up by the YORP effect could result in a binary system from the shedding of mass as centrifugal forces overcome self-gravity and internal strength. Thus, along with spin-up from close planetary encounters (*5–8*) and subcatastrophic collisional fission (*9*), the YORP effect may be responsible for producing a fraction of the near-Earth asteroid binary systems (*10*).

Near-Earth asteroid (54509) 2000 PH5, hereafter referred to as PH5, was discovered by the Massachusetts Institute of Technology Lincoln Laboratory's near-Earth asteroid search program (LINEAR) (*11*) on 3 August 2000 at a geocentric distance of 0.04 astronomical units (AU). PH5 (semimajor axis a = 1.00 AU, eccentricity e = 0.23, inclination i = 1.8°) (fig. S1) is one of only a handful of objects known to be co-orbital companions of Earth (*12–14*). Annual close approaches from 2001 to 2005, as close as 5 lunar distances, were conducive to radar observations and allow us to present a spin-state description and detailed shape model (*15*) of PH5.

We conducted radar observations of PH5 (table S1) using the 70-m antenna and 450 kW, 3.5 cm wavelength transmitter at Goldstone on 27–28 July 2001 (*16*) and the 305-m antenna and 900 kW, 12.6 cm wavelength transmitter at Arecibo on 27–28 July 2004 and 24–26 July 2005. The radar echo is Doppler broadened by the rotation of the target (*17*), and the amount of broadening constrains the spin axis. Daily sums of Doppler-only spectra (fig. S2) determine the radar reflection properties of PH5 (table S2), which are similar to terrestrial planet surfaces (*18*).

Range-Doppler images (*19*) constrain the shape by resolving the radar echo in two orthogonal dimensions: distance from the observer and rotational Doppler shift. High-resolution images of PH5 from Arecibo (Fig. 1, first and fourth columns) with 7.5 m (0.05 µs) resolution reveal an echo 60 to 75 m deep, as well as an array of leading-edge features: convex, linear, and concave, as PH5 rotates. Visual inspection of the range-Doppler images suggests a rotation period of ~12 min, consistent with light-curve observations (*2*). The observing geometries during the 2001 and 2004 observations, which produced the most useful imagery, allow for ~75% surface coverage.

The limb-to-limb bandwidths of Doppler-only spectra (*20*) covering a full rotation of the target produce a bandwidth curve (fig. S3) whose amplitude variation is due to the changing breadth of the rotating nonspherical target on the sky. The mean bandwidth grows from 2001 to 2005, so the line of sight was moving away from the spin axis (*21*). Fits to individual Doppler-only spectra from 2001, 2004, and 2005 with simple ellipsoid and spherical harmonic shape models (*22*) constrain the spin vector to lie within 10° of (180°, –85°) in J2000 ecliptic coordinates at an

Fig. 1. Range-Doppler images (columns 1 and 4) obtained at Arecibo on 28 July 2004, covering one rotation of PH5, along with the corresponding shape model fits to the images (columns 2 and 5), and the plane-of-sky views (columns 3 and 6) of PH5 during the observations. Each 180-m by 180-m frame is separated by ~15° in rotation phase. Radar illumination is from the top of the frame. Range increases from top to bottom, and Doppler frequency increases from left to right; therefore, the rotation of the target appears counterclockwise. Time increases down the left side, then down the right side. The arrow through the plane-of-sky frames indicates the spin vector of PH5.

[1]Department of Astronomy, Cornell University, Ithaca, NY 14853–6801, USA. [2]Institute of Astronomy, Charles University, V Holesovickách 2, 18000 Prague 8, Czech Republic. [3]Department of Aerospace Engineering, University of Michigan, 1320 Beal Avenue, Ann Arbor, MI 48109–2140, USA. [4]Astronomical Institute, Academy of Sciences of the Czech Republic, Fricova 1, CZ-25165 Ondrejov, Czech Republic. [5]School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK. [6]Arecibo Observatory, HC3 Box 53995, Arecibo, PR 00612, USA. [7]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109–8099, USA. [8]University of Maine at Farmington, 173 High Street, Preble Hall, Farmington, ME 04938, USA.

*To whom correspondence should be addressed. E-mail: ptaylor@astro.cornell.edu (P.A.T.); jlm@astro.cornell.edu (J.L.M.)

Fig. 2. Additional rotation phase required to link 20 optical light curves (*2*) from 2001 to 2005 using a shape model with pole (180°, –85°) fit to the 2001 light-curve data. The fitted curve is quadratic in time: $0.5\,\dot{\omega}t^2$, where $\dot{\omega}$ is the rate of change of the spin rate and t is time since the initial epoch of 0^h UT on 27 July 2001. Phases have conservative uncertainties of 10° because of their dependence on the exact shape and orientation of the asteroid.

obliquity of 173° from the orbit normal. This retrograde pole is adopted for shape modeling and translating between the observed light curve and intrinsic sidereal periods in the light-curve analysis (2). A prograde pole fits the observed bandwidths equally well but cannot fit bandwidth curves and light curves simultaneously.

A change in sidereal spin rate is necessary to fit the radar data over time. However, light-curve data alone provide a more accurate determination of the required change than does analysis of the range-Doppler imagery. To produce initial conditions for the spin state of our most detailed shape model, we fit synthetic light curves based on a simple spherical harmonic shape model to the 2001 light-curve data (2), which consist of three epochs over 24 hours, a time frame over which the change in spin rate is negligible. We then include the remaining light-curve data from 2002 to 2005 (2), allowing for an arbitrary phase shift for each light curve to match the phase of the shape model's synthetic light curves. The resulting phase shifts (Fig. 2) necessary to link the light curves are well fit by a quadratic function in time; in other words, the spin rate is increasing linearly with time. The use of a linear change in the spin rate rather than a constant spin rate yields improvement by a factor of ten in the reduced χ^2 value of a spherical harmonic fit to the entire collection of light-curve data.

For pole coordinates of (180°, –85°), the sidereal spin rate necessary to fit the light curve data are 42582.41 ± 0.02 deg/day (12.17-min period) at the initial epoch of 0^h UT on 27 July 2001, and the necessary change in spin rate is $(2.0 \pm 0.2) \times 10^{-4}$ deg/day^2. This determination of a continuous increase in spin rate precisely matches the discrete spin-period changes observed in (2). The fractional change in spin period is -1.72×10^{-6} ($\pm 10\%$) per year. Accelerations determined by this method for poles less than 10° from (180°, –85°) lie within 8% of the nominal acceleration, indicating relative insensitivity of the acceleration to precise pole location.

The shape models (22) of PH5 are 288-vertex polyhedra with 572 triangular facets giving the models 12° resolution in longitude, twice the amount PH5 rotated by during the exposure time for each range-Doppler image. We produced a family of models with a range of surface "roughness" based upon large-, medium-, and small- scale topography (22) to determine both the shapes that best fit the combined radar and light-curve data and how roughness affects the YORP acceleration calculations. All models produced have similar silhouettes, with much of the variation coming from the smoothness of the surface and the length of the shortest principal axis of inertia. The best PH5 shape model shows very good agreement with all range-Doppler images, reproducing the various convex, linear, and concave leading edges of the echoes (Fig. 1). The shape of PH5 (fig. S4) is distinguished by its flattened northern hemisphere with a linear edge

Table 1. Summary of shape models and predictions for the change in spin period due to the YORP effect. Shape A (rough) is the best-fit shape model depicted in Fig. 1, and the other shape models sample the 1-σ formal uncertainty region for the fit. "Smooth" and "rough" describe the amount of facet-scale topography the shape models allow. *a*, *b*, and *c* are the extents of the shape model along the principal axes of inertia. *D* is the diameter of a sphere with the same volume as the shape model. The YORP predictions are given by the factor by which they overestimate the observed change in spin period. Shape models with smoother surfaces result in predicted changes in spin period more consistent with observation than do rough surfaces, which provide better fits to the radar and light-curve data. Several factors could account for the discrepancy between observed and predicted values.

Shape	Type	*a/b/c* (m)	*D* (m)	$\Delta P/P_0$ per year factor	
				Model 1 (3)	Model 2 (26)
A	Smooth	150/128/93	114.2	2.86	3.72
	Rough	149/134/96	112.8	6.18	6.03
B	Smooth	149/130/91	113.2	2.95	3.36
	Rough	147/132/91	111.7	4.56	4.89
C	Smooth	149/129/97	115.1	2.63	3.24
	Rough	149/131/99	113.0	4.74	5.14
			Observed $\Delta P/P_0$ per year (2) = -1.72×10^{-6}		

and prominent concavity that are clearly visible in the radar images.

The phase agreement between the rotating model and the entire set of range-Doppler images, as well as the ability to link the light curves from the 4-year optical photometry campaign (2), is due to the inclusion of a linear change in the intrinsic spin rate of PH5 (23). Harder to reproduce is the large amplitude of the PH5 light curves (fig. S5). The discrepancy may be due to unresolved topography, shadowing effects from concavities, albedo variations, or deficiencies in the photometric model used in the shape modeling software.

The rapid rotation of PH5, the increasing spin rate, and the near-180° obliquity are consistent with simulations of a body subjected to YORP torques (3). YORP torques can change the spin state of PH5 on less than million-year time scales (2), shorter than the dynamical (2, 24) and collisional lifetimes (25) of about 10 million and 1 billion years, respectively, that would reorient the spin vector or disrupt the asteroid. Despite the repeated close encounters between PH5 and Earth, planetary tidal torques are not strong enough (2) to cause the observed change in spin rate. Without other plausible causes, the YORP effect is the most viable mechanism for explaining the observations.

Using the PH5 spin state and shape models, two independent YORP acceleration models (3, 26) predict changes in spin rate, in terms of the fractional change in spin period ($\Delta P/P_0$) per year, 2 to 7 times as large as those observed (Table 1). Smooth models, those with less facet-scale topography, produce changes in spin rate closer by a factor of 2 to the observed value as models with rougher surfaces. Several factors may account for the discrepancy between observed and theoretical values. The incomplete surface coverage by radar of PH5's irregular shape undoubtedly affects the accuracy of the shape model. Mismodeled morphology results in erroneous YORP torque values

and accelerations that could cause errors even greater than the factor of 2 seen between smooth and rough models. Unresolved surface characteristics much finer than the radar resolution may also be very important for accurate torque calculations. Furthermore, we assume a bulk density ρ of 2.5 g/cm^3. Because $\Delta P/P_0$ per year scales as $(\rho D^2)^{-1}$, a combination of higher density and larger size could reduce the discrepancy. Deficiencies in the YORP simulations, including the thermal model formulation and uncertainties in key thermal parameters and surface-scattering properties, may also play a role. Simulations assume an ideal Lambertian scattering surface; in general, a non-ideal surface would produce a result more consistent with observation. Although PH5 has significantly nonzero thermal inertia (27), trials of surface thermal conductivities between 0.005 and 0.05 W/m/K show no appreciable change in results, as expected (3, 26). Altogether, a combination of incomplete surface coverage and assumptions about density, thermal parameters, and scattering properties are likely responsible for the discrepancy we find. The order-of-magnitude agreement between observation and theory provides the best evidence to date that the YORP effect acts upon small solar-system bodies.

The observation of a continuously increasing asteroid spin rate and the reasonable agreement with theoretical models support the YORP effect as an important process for altering the spin state of small solar system bodies. YORP may therefore explain several puzzling issues in asteroid dynamics, including the observed spin-rate distribution of asteroids <10 km in diameter and the production of binary systems.

References and Notes
1. D. P. Rubincam, *Icarus* **148**, 2 (2000).
2. S. C. Lowry *et al.*, *Science* **316**, 272 (2007); published online 8 March 2007; 10.1126/science.1139040.
3. D. Čapek, D. Vokrouhlický, *Icarus* **172**, 526 (2004).
4. P. Pravec, A. W. Harris, *Icarus* **148**, 12 (2000).
5. W. F. Bottke, H. J. Melosh, *Nature* **381**, 51 (1996).

6. E. Asphaug, W. Benz, *Icarus* **121**, 225 (1996).
7. D. C. Richardson, W. F. Bottke, S. G. Love, *Icarus* **134**, 47 (1998).
8. J. L. Margot *et al.*, *Science* **296**, 1445 (2002).
9. S. J. Weidenschilling, P. Paolicchi, V. Zappalà, in *Asteroids II*, R. P. Binzel, T. Gehrels, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, AZ, 1989), pp. 643–658.
10. W. F. Bottke, D. Vokrouhlický, D. P. Rubincam, M. Broz, in *Asteroids III*, W. F. Bottke, P. Paolicchi, R. P. Binzel, A. Cellino, Eds. (Univ. of Arizona Press, Tucson, AZ, 2002), pp. 395–408.
11. G. H. Stokes, J. B. Evans, H. E. M. Viggh, F. C. Shelly, E. C. Pearce, *Icarus* **148**, 21 (2000).
12. P. Wiegert *et al.*, American Geophysical Union, Fall Meeting 2002, abstr. #P11A-0352 (2002).
13. J. L. Margot, P. D. Nicholson, *AAS/Division of Dynamical Astronomy Meeting* **34** (2003).
14. R. Brasser *et al.*, *Icarus* **171**, 102 (2004).
15. S. Hudson, *Remote Sensing Rev.* **8**, 195 (1993).
16. J. L. Margot *et al.*, *Bull. Am. Astron. Soc.* **35**, 960 (2003).
17. The Doppler broadening of the radar echo due to rotation of the target is $B = (4\pi D/\lambda P)\sin\alpha$, where B is the limb-to-limb bandwidth of the echo, D is the target diameter producing the Doppler shift at the current viewing geometry and rotation phase, λ is the radar wavelength, P is the spin period of the target, and α is the inclination of the spin axis to the line of sight.
18. S. J. Ostro, *Rev. Mod. Phys.* **65**, 1235 (1993).
19. Resolution in time delay, and equivalently range, is achieved by transmitting a time-dependent signal and analyzing the received signal according to arrival time.

The time increment τ used in the transmitted signal yields a range resolution $c\tau/2$, where c is the speed of light.
20. We typically define the limb-to-limb bandwidth as the full width of the radar echo at the level of twice the root mean square (RMS) of the off-DC, off-target noise. The exception is the strong 2004 Arecibo data, for which we use 10 times the RMS as the threshold to avoid contributions from frequency sidelobes.
21. This assumes PH5 is a principal axis (PA) rotator where the spin axis remains fixed in inertial space and aligned with the axis of maximum moment of inertia. The spin axis of PH5 must then be oriented such that the angles it makes with the lines of sight satisfy the observed bandwidths (*17*). The damping time scale (*28*) to PA rotation for PH5 is of order 0.1 million years.
22. Materials and methods are available as supporting material on *Science* Online.
23. The spin state solution is also validated by the phase agreement of infrared lightcurves from the Spitzer Space Telescope with synthetic lightcurves produced with our shape (*27*).
24. B. Gladman *et al.*, *Science* **277**, 197 (1997).
25. W. F. Bottke Jr., M. C. Nolan, R. Greenberg, R. A. Kolvoord, in *Hazards Due to Comets and Asteroids*, T. Gehrels, M. S. Matthews, A. M. Schumann, Eds. (Univ. of Arizona Press, Tucson, AZ, 1994), pp. 337–357.
26. D. J. Scheeres, *Icarus* 10.1016/j.icarus.2006.12.015 (2007).
27. M. Mueller, A. W. Harris, *IAU General Assembly Abstracts* (2006), p. 95.
28. I. Sharma, J. A. Burns, C.-Y. Hui, *Mon. Not. R. Astron. Soc.* **359**, 79 (2005).
29. We thank the staffs of the Arecibo Observatory and the Goldstone Solar System Radar for their support in performing this research. The Arecibo Observatory is part of the National Astronomy and Ionosphere Center, which is operated by Cornell University under a cooperative agreement with NSF. Some of this work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. This material is based in part on work supported by NASA under the Science Mission Directorate Research and Analysis Programs. P.A.T. and J.L.M. were partially supported by NASA grant NNG04GN31G. The work of P.P. and D.V. was supported by the Grant Agency of the Czech Republic. D.J.S. acknowledges support from the NASA Planetary Geology and Geophysics Program. S.C.L. and A.F. acknowledge support from the Leverhulme Trust and PPARC, respectively. C.M. was partially supported by NSF grant AST-0205975. The International Astronomical Union has approved the name YORP for asteroid (54509) 2000 PH5.

Analyses of Soft Tissue from *Tyrannosaurus rex* Suggest the Presence of Protein

Mary Higby Schweitzer,[1,2,3]* Zhiyong Suo,[4] Recep Avci,[4] John M. Asara,[5,6] Mark A. Allen,[7] Fernando Teran Arce,[4,8] John R. Horner[3]

We performed multiple analyses of *Tyrannosaurus rex* (specimen MOR 1125) fibrous cortical and medullary tissues remaining after demineralization. The results indicate that collagen I, the main organic component of bone, has been preserved in low concentrations in these tissues. The findings were independently confirmed by mass spectrometry. We propose a possible chemical pathway that may contribute to this preservation. The presence of endogenous protein in dinosaur bone may validate hypotheses about evolutionary relationships, rates, and patterns of molecular change and degradation, as well as the chemical stability of molecules over time.

It has long been assumed that the process of fossilization results in the destruction of virtually all original organic components of an organism, and it has been hypothesized that original molecules will be either lost or altered to the point of nonrecognition over relatively short time spans (well under a million years) (*1–7*). However, the discovery of intact structures retaining original transparency, flexibility, and other characteristics in specimens dating at least to the Cretaceous (*8, 9*) suggested that, under certain conditions, remnant organic constituents may persist across geological time.

The skull, vertebrae, both femora and tibiae, and other elements of an exceptionally well-preserved *Tyrannosaurus rex* [MOR 1125 (*8*)] were recovered from the base of the Hell Creek Formation in eastern Montana (USA), buried within at least 1000 m³ of medium-grained, loosely consolidated sandstone interfingered with fine-grained muds, interpreted as stream channel sediments. Demineralization of femur and tibia fragments revealed the preservation of fibrous, flexible, and apparently original tissues, as well as apparent cells and blood vessels (*8*), but the endogeneity and composition of these structures could not be ascertained without further analyses.

We present molecular and chemical (*10*) analyses of tissues remaining after partial demineralization (*11*) of the left and right femora and associated medullary bone (*12*) that would, in extant bone, represent the extracellular matrix (osteoid) dominated by collagen I (*13*). Because

of its ordered structure as a triple helix (*14, 15*), collagen I has unique characteristics that are highly conserved across taxa, making validation of its presence relatively straightforward. The molecular composition of collagen incorporates glycine, the smallest amino acid, at every helical turn. Therefore, an amino acid profile of collagen results in ~33% glycine content (*14*). This molecular structure also results in packing of microfibrils with a banded repeat of ~70 nm (*15, 16*). Collagen also shows posttranslational hydroxylation of about half of all proline and some lysine residues; thus, the detection of hydroxyproline and hydroxylysine in extracts of organic material is viewed as strong evidence for the presence of collagen (*17, 18*). Finally, collagen is identified by polyclonal or monoclonal antibody reactivity that can distinguish between collagen types (*19*). We focused on identifying collagen-like compounds because in addition to being abundant and easily identified by multiple

[1]Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, USA. [2]North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA. [3]Museum of the Rockies, Montana State University, Bozeman, MT 59717, USA. [4]Image and Chemical Analysis Laboratory Facility, Department of Physics, Montana State University, Bozeman, MT 59717, USA. [5]Division of Signal Transduction, Beth Israel Deaconess Medical Center, Boston, MA 02115, USA. [6]Department of Pathology, Harvard Medical School, Boston, MA 02115, USA. [7]Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT 59717, USA. [8]Center for Nanomedicine, Pulmonary and Critical Care Medicine, Department of Medicine, University of Chicago, Chicago, IL 60637, USA.

*To whom correspondence should be addressed. E-mail: schweitzer@ncsu.edu

and independent methods, this protein is durable (*20, 21*) and resistant to degradation.

The fibrous nature of demineralized dinosaur tissues was demonstrated by optical (*8*) and electron (fig. S1) microscopy. Furthermore, regions of dinosaur cortical and medullary (*12*) bone demonstrated a repeat pattern with periodicity of ~70 nm when examined by atomic force microscopy (AFM) (Fig. 1, A to D), consistent with collagen in extant bone (Fig. 1, E and F) and similar to that previously observed in fragments of demineralized Cretaceous avian bone (*22*). However, this periodic pattern was rarely observed in ultrathin sections of MOR 1125 demineralized bone by transmission electron microscopy (TEM) (fig. S1). This may be a methodological problem, or the periodic features we observe (Fig. 1, A to D) may be due to surface features generated when demineralization removed most of the apatite crystals emplaced during biomineralization, when collagen acted as a template. Thus, the banded features may represent a type of natural molecular imprinting (*23*), because banded fibers have been observed by TEM for other dinosaur tissues (*24*).

TEM studies confirm that, unlike extant bone, dinosaur bone did not completely demineralize after prolonged incubation in EDTA (*11*). Selected-area electron diffraction (SAED) of the tissues (fig. S1D, inset) showed that this retained mineral is biogenic hydroxylapatite (*25*). It is not possible to determine this conclusively because of the similarity in structure between hydroxylapatite and fluorapatite; however, the observed diffraction circle intensities are most consistent with hydroxylapatite. This finding suggests that the bone mineral is virtually unchanged from the living state and has undergone little if any alteration.

Force curve measurements of demineralized dinosaur medullary and cortical bone indicate that the elasticity of dinosaur tissues was similar to that of demineralized extant bone. We measured both embedded sections (fig. S2A) and unembedded whole mounts (fig. S2B) of demineralized bone in both air and liquid (*11*). The demineralized bone surface softened after exposure to buffer, allowing the AFM tip to penetrate deeper into the tissues with less resistance. Thus, the modulus of elasticity (fig. S2C) was reduced in liquid by more than three orders of magnitude (fig. S2B). Although ~2000 nN of force was required to penetrate ~40 nm into MOR 1125 bone matrix in air, only ~15 nN of force was required to depress the tip ~75 nm into the same matrix when hydrated (fig. S2B, inset).

MOR 1125 cortical and medullary whole-bone extracts showed reactivity to antibodies raised against chicken collagen I (*11*) when measured by enzyme-linked immunosorbent assay (ELISA), although the degree of binding varied widely. Reactivity was greatly reduced in dinosaur extracts relative to extant samples (fig. S3), but still at least twice that observed in negative con-
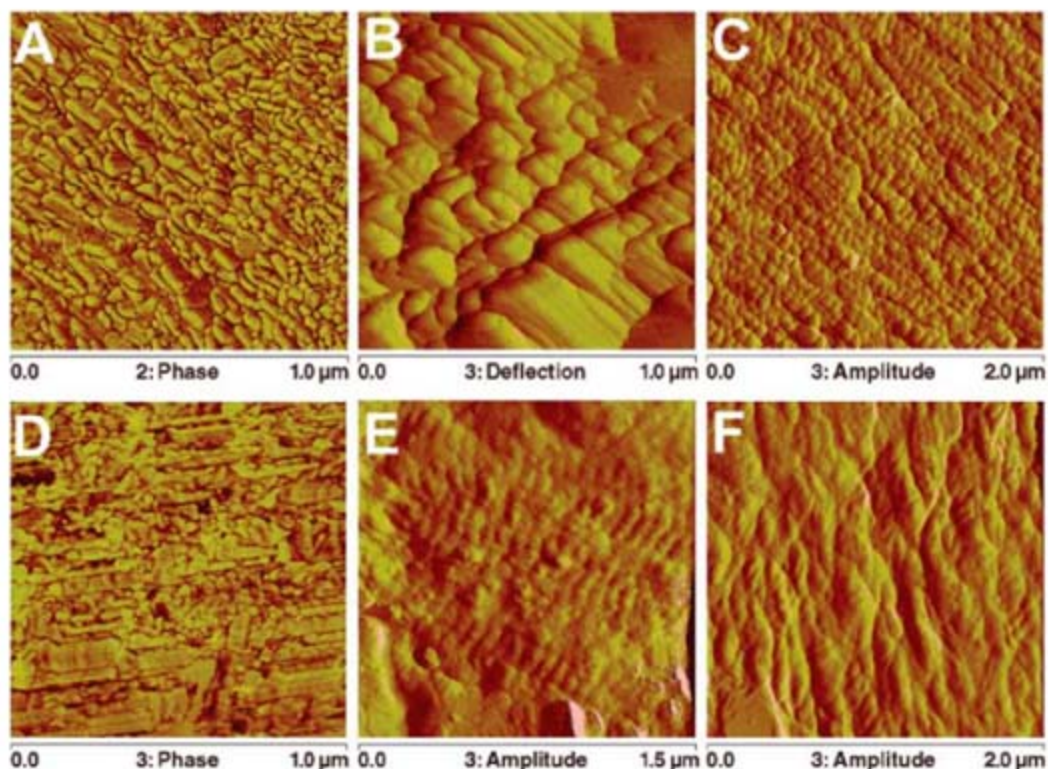
Fig. 1. AFM images of partially demineralized bones of MOR 1125 (**A** to **D**) and emu (**E** and **F**). (A) Phase image of MOR 1125 cortical bone imaged in air; (B) deflection image of MOR cortical bone imaged in phosphate-buffered saline; (C) amplitude image of embedded and sectioned MOR medullary bone imaged in air; (D) phase image of MOR 1125 medullary bone imaged in air (note longitudinal and cross-sectional orientation of fiber-like structures at right angles to each other); (E) amplitude image of emu cortical bone imaged in air; (F) amplitude image of emu medullary bone imaged in air.

Fig. 2. In situ immunochemistry on 300-nm sections of demineralized MOR 1125 cortical bone (**A** to **D**) and medullary bone (**E** to **H**). (A) and (E), no primary antibodies added (negative control); (B) and (F), antibodies to avian collagen I; (C) and (G), antibodies to actin protein (nonrelevant, negative control); (D) and (H), antibodies to avian collagen I, inhibited by incubating with purified chicken collagen before exposing to dinosaur tissues. All data were collected using the same parameters at 122-ms integration. (**I** to **K**) MOR 1125 cortical tissue exposed to (I) no primary, (J) antibodies to avian collagen I, or (K) collagenase digestion followed by antibodies to avian collagen I, as described (*11*). Data in (I), (J), and (K) were collected at 149-ms integration.
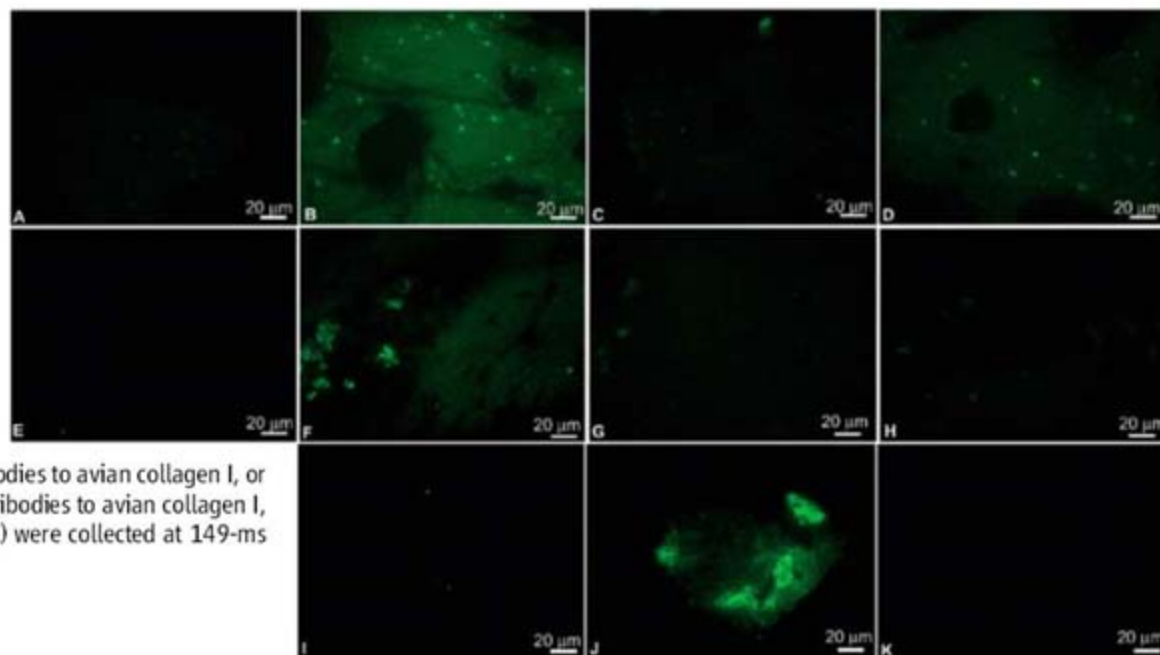
trols of coextracted sediments and buffer without sample, similarly treated.

We confirmed the antibody reactivity data by in situ immunohistochemistry in a series of experiments. We exposed thin (0.3 to 0.5 μm) sections of demineralized cortical (Fig. 2, A to D and I to K) and medullary (Fig. 2, E to H) dinosaur bone to antibodies raised against avian collagen I, both before (Fig. 2, B and F) and after inhibition of antibodies with chicken collagen (Fig. 2, D and H) (11). Additionally, antibody reactivity (Fig. 2J) was significantly decreased after we digested dinosaur tissues with collagenase (Fig. 2K), although this enzyme effect was not consistently observed. Reactivity to antibodies, measured by fluorescence, was significantly greater than in negative controls (Fig. 2, A, C, E, G, and I) and was localized to tissues. We also observed some binding of osteocalcin antibodies to dinosaur tissues (fig. S4). These patterns were similar to those observed with extant emu cortical and medullary bone (fig. S5). Immunoreactivity in dinosaur tissues was greatly reduced from that observed in extant bone, as illustrated by longer integration times and fainter signal, but was greater than in negative controls. Immunohistochemistry performed on sediments was negative for binding. These results imply that the concentration of reactive epitopes in the dinosaur

tissues is very low, consistent with the ELISA results. That antibody reactivity was more consistently observed in situ than in ELISA could be due to greater alteration and/or loss of organic compounds during extraction procedures, or to reduced binding of degraded antigen to ELISA plate polymers.

The presence of collagen-derived epitopes in demineralized tissues is supported by mass spectrometry data. Time-of-flight secondary ion mass spectrometry (TOF-SIMS) detects surface ions associated with molecular fragmentation with high mass resolution, and can localize signal to whole samples without subjecting them to chemical extraction. In situ TOF-SIMS analyses were performed to unambiguously detect amino acid residues consistent with the presence of protein in demineralized MOR 1125 tissues (Fig. 2 and fig. S6). We obtained ratios of glycine (Gly), the most abundant amino acid in collagen [~33% (14)], and alanine (Ala), which constitutes about 10% of collagenous amino acids, to support the presence of the specific collagen α1 type 1 protein in these tissues. Small peaks representing proline (Pro) at mass/charge ratio (m/z) 70 (Fig. 3C), lysine (Lys) at m/z 84 (fig. S6A), and leucine or isoleucine at m/z 86 (fig. S6B) were also detected. TOF-SIMS is highly matrix dependent, and desorption and ionization of some amino acid res-

idues, especially modified residues such as hydroxylated Pro, are less efficient than for other residues (26). These modified residues were not detected by this method but were readily identified by other mass spectrometry methods (10).

The Gly:Ala ratio for published chicken collagen α1 type 1 sequence (27) is 2.5:1. The TOF-SIMS results show that the Gly:Ala ratio in medullary bone of MOR 1125 is 2.6:1 (Fig. 3, A and B). Sandstones entombing the dinosaur, subjected to TOF-SIMS as a control, showed little or no evidence for these amino acids (Fig. 3, D and E).

We identified a variety of nitrogen-containing species—including an alkyl amine group, $C_7H_{18}N_2^+$, located at 130 amu (fig. S6C)—in all dinosaur samples tested, but not in any surrounding sediments. We also observed a number of Fe-C-H species such as $FeCH$, $FeCH_2$, and $FeCH_3$, associated with the dinosaur matrix (fig. S7) but not seen in extant material. Similar compounds were observed in the sediments surrounding the dinosaur. These may be microbial products, as sequences from iron-containing microbial enzymes were identified by mass spectrometry (10). We interpret these fragments as evidence that iron may help preserve soft tissue through initiation of intra- and intermolecular cross-links (9).

Dinosaur protein sequence, including collagen, should be most similar to that of birds among extant taxa, according to other phylogenetic information (28). The hypothesis that molecular fragments of original proteins are preserved in the mineralized matrix of bony elements of MOR 1125 is supported by peptide sequences recovered from dinosaur extracts, some of which align uniquely with chicken collagen α1 type 1 (10).

The amount of protein or protein-like components in MOR 1125 is minimal. The percent yield after extraction and lyophilization was ~0.62% for cortical bone and 1.3% for medullary bone. Protein-derived material is only a small percentage of the lyophilate relative to other material coextracted from bone, as assessed by comparison of immunoreactivity with extant samples. This is verified by mass spectrometry, which identifies only femtomole amounts of sequenceable material (10) in a heterogeneous mixture of extracted material.

Microenvironments within a single bone vary greatly, and not every fragment of bone examined yielded positive results. There was a high degree of variability between extractions, and we have also noted progressive reduction of signal in more recent extractions, indicating bone degradation in modern environments (29). Therefore, each of the analyses we report has been repeated numerous times, and we have set a minimum of three repetitions with similar results before reporting an assay as positive. Additionally, experiments have been conducted independently in at least three different labs and by numerous investigators, and the results strongly support the endogenicity of collagen-like protein molecules.

We hypothesize that these molecular fragments are preserved because reactive sites on the
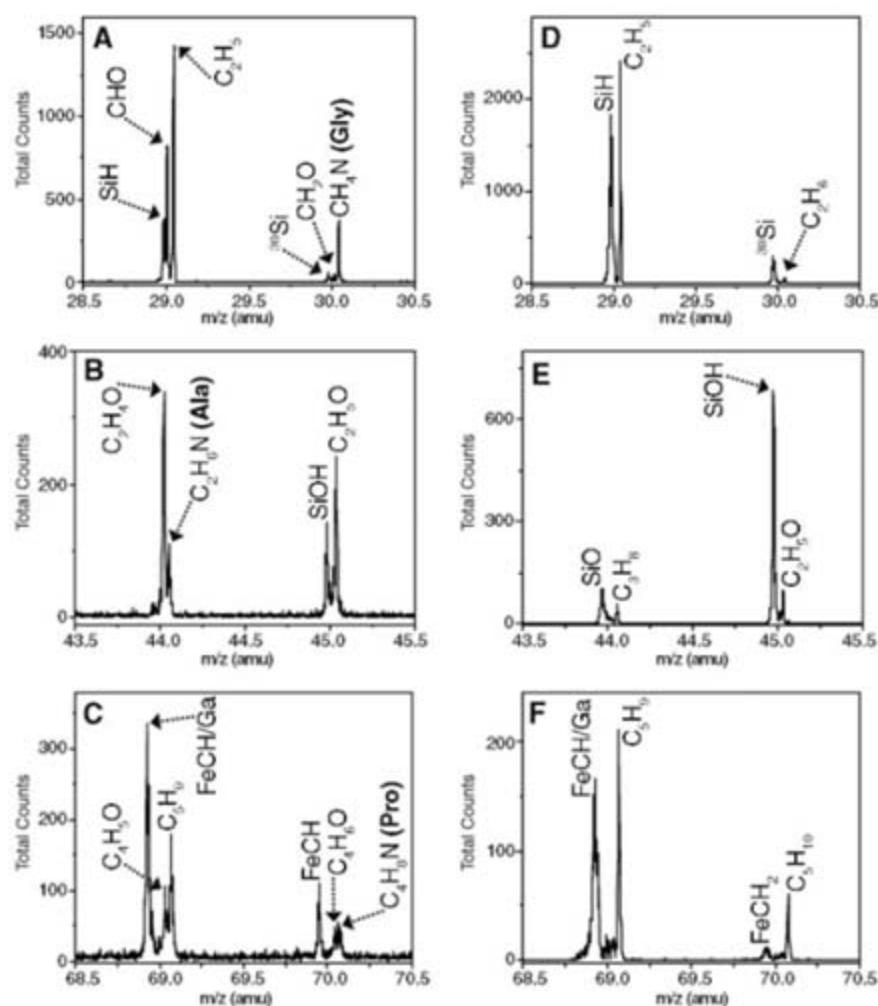
Fig. 3. TOF-SIMS spectra of demineralized MOR 1125 medullary bone (**A** to **C**) and entombing sedimentary matrix (**D** to **F**). The imonium ions for Gly ($m/z = 30$), Ala ($m/z = 44$), and Pro ($m/z = 70$) can be unambiguously identified for MOR 1125; no signal was observed in sediment controls that corresponded to these amino acids. See text for discussion.

original protein molecules became irreversibly cross-linked, both to similar molecules and to mineral or exogenous organic components. These cross-linking reactions may have been initiated by unstable metal ions that formed free radicals (30, 31), which in turn reacted with organic molecules to form polymers (6, 7, 9, 32). We propose that the unstable metal ions were derived from the post mortem degradation of iron-containing dinosaur biomolecules such as hemoglobin, myoglobin, and possibly cytochromes (9, 31). Once stabilized by these cross-linking reactions, the molecules were no longer available as substrates for further degradative reactions.

The intimate relationship between apatite and the organic phase of bone also contributes to the preservation of organic matter (16, 33–38), but we propose that the mineral phase may be stabilized by this relationship as well. The presence of biogenic apatite in these 68-million-year-old bones can only be rationalized by protection from an intact organic phase, which in turn is only satisfied by a synergistic relationship between collagen and mineral phases. Whereas extant bone retains no detectable calcium after days to weeks of demineralization, dinosaur bone retains a fraction of recognizable apatite crystals after months of treatment (fig. S1). Another contributing factor in the retention of original mineral may be that apatite is stabilized in the presence of calcite (33). Sandstones surrounding MOR 1125 contain abundant calcite cements.

The depositional environments may affect organic preservation in other ways. Comparison of fossils from a variety of environments indicates that those derived from sandstones are more likely to retain soft tissues and/or cells (9). We hypothesize that the porosity of sandstones may facilitate draining of enzymes of decay and suppurating fluids as the organism degrades, whereas organisms buried in nonporous mudstones or clays may be exposed to these longer and therefore may be more completely degraded.

Our findings indicate the need for optimizing methods of extraction and handling of fossil material. In particular, the decrease in signal we observed over time supports the need to establish field collection and storage of fossils according to protocols that allow future analytical studies (29).

The data presented here illustrate the value of a multidisciplinary approach to the characterization of very old fossil material and validate sequence data reported elsewhere (10). The inclusion of fossil-derived molecular sequences into existing phylogenies may provide greater resolution and may allow reconstruction of character evolution beyond what is currently possible. Elucidating modifications to ancient molecules may shed light on patterns of degradation and diagenesis. The presence of original molecular components is not predicted for fossils older than a million years (1–7), and the discovery of collagen in this well-preserved dinosaur supports the use of actualistic conditions to formulate molecular degradation rates and models, rather than relying on theoretical or experimental extrapolations derived from conditions that do not occur in nature.

References and Notes

1. T. Lindahl, *Nature* **365**, 700 (1993).
2. G. Eglinton, G. A. Logan, *Philos. Trans. R. Soc. London Ser. B* **333**, 315 (1991).
3. J. L. Bada, X. Y. S. Wang, H. Hamilton, *Philos. Trans. R. Soc. London Ser. B* **354**, 77 (1999).
4. M. Hoss, *Nature* **404**, 453 (2000).
5. D. E. G. Briggs, R. P. Evershed, M. J. Lockheart, *Paleobiology* **26**, 169 (2000).
6. D. E. G. Briggs, *Annu. Rev. Earth Planet. Sci.* **31**, 275 (2003).
7. B. A. Stankiewicz et al., *Geology* **28**, 559 (2000).
8. M. H. Schweitzer, J. L. Wittmeyer, J. R. Horner, J. B. Toporski, *Science* **307**, 1952 (2005).
9. M. H. Schweitzer, J. L. Wittmeyer, J. R. Horner, *Proc. R. Soc. London Ser. B* **274**, 183 (2007).
10. J. M. Asara, M. H. Schweitzer, L. M. Freimark, M. Phillips, L. C. Cantley, *Science* **316**, 280 (2007).
11. See supporting material on *Science* Online.
12. M. H. Schweitzer, J. L. Wittmeyer, J. R. Horner, *Science* **308**, 1456 (2005).
13. L. Zylberberg, *C. R. Palevol.* **3**, 591 (2004) and references therein.
14. M. van der Rest, *Bone* **3**, 187 (1991).
15. V. Ottani, M. Raspanti, A. Ruggeri, *Micron* **32**, 251 (2001).
16. S. Weiner, H. D. Wagner, *Annu. Rev. Mater. Sci.* **28**, 271 (1998).
17. G. Nemethy, H. A. Scherega, *Biochemistry* **25**, 3184 (1986).
18. J. G. Bann, H. P. Bachinger, *J. Biol. Chem.* **275**, 24466 (2000).
19. D. R. Keene, L. Y. Sakai, R. E. Burgeson, *J. Histochem. Cytochem.* **39**, 59 (1991).
20. N. Tuross, L. Stathoplos, *Methods Enzymol.* **224**, 121 (1993).
21. P. Semal, R. Orban, *J. Archaeol. Sci.* **22**, 463 (1995).
22. R. Avci et al., *Langmuir* **21**, 3584 (2005).
23. K. Mosbach, *Trends Biochem. Sci.* **19**, 9 (1994).
24. F. A. Rimblot-Baly, A. de Ricqles, L. Zylberberg, *Ann. Paleontol.* **81**, 49 (1995).
25. J. M. Hughes, M. Cameron, K. D. Crowley, *Am. Mineral.* **74**, 870 (1989).
26. B.-A. Gotliv et al., *J. Struct. Biol.* **156**, 320 (2006).
27. Accession number P02457, NCBI nonredundant protein database.
28. T. Holtz, *Gaia* **15**, 5 (1998) and references therein.
29. M. Pruvost et al., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 739 (2007).
30. J. M. C. Gutteridge, D. A. Rowley, B. Halliwell, *Biochem. J.* **199**, 263 (1981).
31. P. Mladěnka, T. Simunek, M. Hubl, R. Hrdina, *Free Radic. Res.* **40**, 263 (2006).
32. F. Q. Schafer, S. Yue Qian, G. R. Buettner, *Cell. Mol. Biol.* **46**, 657 (2000).
33. F. Berna, A. Matthews, S. Weiner, *J. Archaeol. Sci.* **31**, 867 (2004).
34. M. J. DeNiro, S. Weiner, *Geochim. Cosmochim. Acta* **52**, 2415 (1988).
35. M. J. Glimcher, L. Cohen-Solal, D. Kossiva, A. deRicqles, *Paleobiology* **16**, 219 (1990).
36. T. Schmidt-Schultz, M. Schultz, *Am. J. Phys. Anthropol.* **123**, 30 (2004).
37. G. A. Sykes, M. J. Collins, D. I. Walton, *Org. Geochem.* **23**, 1059 (1995).
38. M. Salamon, N. Tuross, B. Arensburg, S. Weiner, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13783 (2005).
39. We thank J. Wittmeyer for sample preparation and data collection for many of our analyses; N. Blair, S. Brumfield, N. Equall, B. Glaspey, L. Kellerman, J. Monds, R. Mecham, and M. Tientze; M. Franklin, C. Paden, R. Wilkinson, and J. Starkey for lab facilities; M. Dykstra, J. Phillips, and W. Savage for imaging; W. Zheng for supporting data; and the Museum of the Rockies field crew responsible for the recovery of MOR 1125, "B. rex." Supported by NSF grants EAR-0541744 and EAR-0548847 and the David and Lucile Packard Foundation (M.H.S.), NASA Experimental Program to Stimulate Competitive Research grant NCC5-579 (R.A.), NSF grant EAR-0634136 (J.M.A.), and N. Myrhvold (J.R.H.).

Protein Sequences from Mastodon and *Tyrannosaurus Rex* Revealed by Mass Spectrometry

John M. Asara,[1,2]* Mary H. Schweitzer,[3] Lisa M. Freimark,[1] Matthew Phillips,[1] Lewis C. Cantley[1,4]

Fossilized bones from extinct taxa harbor the potential for obtaining protein or DNA sequences that could reveal evolutionary links to extant species. We used mass spectrometry to obtain protein sequences from bones of a 160,000- to 600,000-year-old extinct mastodon (*Mammut americanum*) and a 68-million-year-old dinosaur (*Tyrannosaurus rex*). The presence of *T. rex* sequences indicates that their peptide bonds were remarkably stable. Mass spectrometry can thus be used to determine unique sequences from ancient organisms from peptide fragmentation patterns, a valuable tool to study the evolution and adaptation of ancient taxa from which genomic sequences are unlikely to be obtained.

Obtaining genome sequences from a number of taxa has dramatically enhanced our abilities to study the evolution and adaptation of organisms. However, difficulties in the acquisition of DNA or RNA from ancient extinct taxa limit the ability to examine molecular evolution. Recent advances in mass spectrometry (MS) technologies have

made it possible to obtain sequence information from subpicomolar quantities of fragmented proteins and peptides (1, 2), but the conversion of these fragmentation patterns (MS/MS spectra) into peptide sequences in the absence of genomic and protein sequences from publicly available databases has been a challenge. If the unknown peptide is identical in sequence to a protein region from an organism whose genes or proteins have previously been sequenced, then the fragmentation pattern (the mass/charge ratios

and relative intensities of peaks) will match a theoretical fragmentation pattern from a sequence in publicly available protein databases or from the fragmentation pattern of a synthetically derived peptide, to confirm its identity.

Using this approach, tryptic peptide sequences of collagen have been identified from a 100,000- to 300,000-year-old mammoth skull, and these matched collagen fragments of extant mammalian taxa including bovine, a result that was also supported by immunological methods (3). MS has also been used to report protein sequences from younger fossil specimens (4–6). However, sequence data from very old (1 million years or older) fossils has been hindered by protein concentrations below the limits of detection by most analytical methods, and by theoretical limits based on predicted rates of degradation (7, 8). In addition, most commercial software for identifying peptide sequences by MS relies on the peptide fragmentation pattern matching identically to that of a peptide/protein sequence in existing sequence

databases. Here we show that these hindrances can be overcome by a two-step proteomics approach to obtain sequences from ion-trap MS fragmentation patterns.

We sequenced collagen protein fragments derived from fossilized bones of two extinct taxa: a 160,000- to 600,000-year-old mastodon [specimen number Museum of the Rockies (MOR) 605] (9) and a 68-million-year-old dinosaur (Tyrannosaurus rex, MOR 1125) (10), results that are supported by immunological and molecular analyses published in this issue by Schweitzer et al. (11). We first looked for tryptic peptide fragments from extracts of fossilized bone that matched identically with sequences from an orthologous protein or proteins from extant taxa, thereby identifying the protein(s) of interest. This is a common procedure for conserved proteins from taxa that share genomic information. Next, we generated a protein sequence database of likely drifts in amino acids in other tryptic peptides by comparing amino acid sequences of the orthologs

[1]Division of Signal Transduction, Beth Israel Deaconess Medical Center, Boston, MA 02115, USA. [2]Department of Pathology, Harvard Medical School, Boston, MA 02115, USA; North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA; and Museum of the Rockies, Montana State University, Bozeman, MT 59717, USA. [3]Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, USA. [4]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: jasara@bidmc.harvard.edu
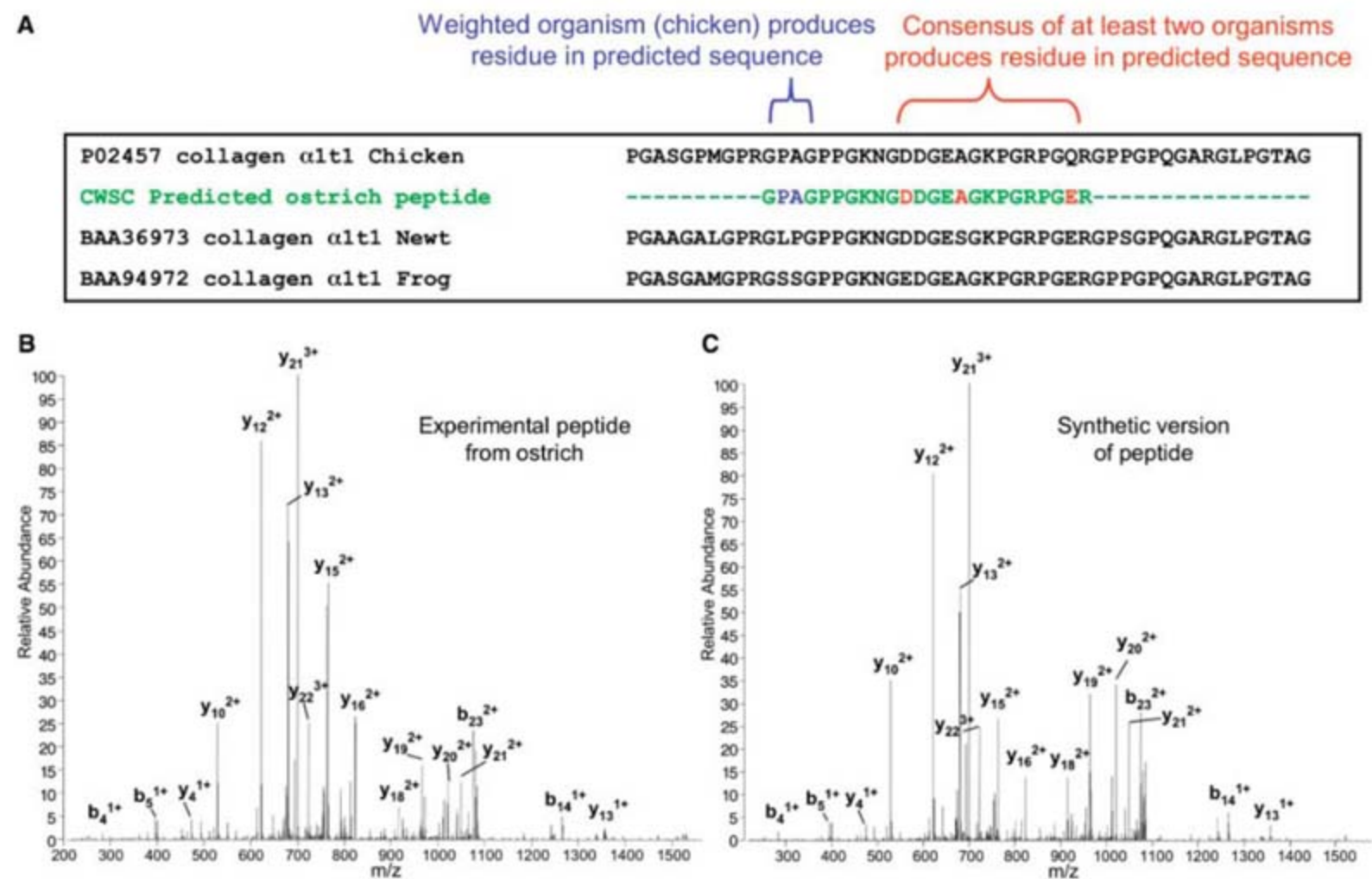
Fig. 1. Sequence identification by matching a peptide fragmentation pattern to a predicted tryptic peptide sequence. (**A**) Example of the chicken weighted simple consensus (CWSC) sequence algorithm for predicting a tryptic collagen peptide sequence from a previously unsequenced taxon (ostrich) based on three related organisms with one weighted organism (chicken). If a consensus of at least two organisms was present at amino acid residues that diverged (positions 11, 15, and 23), the consensus residue was chosen in the predicted tryptic peptide sequence. For residues that diverged where no consensus was present (positions 2 and 3), the residue from the weighted organism (chicken) was chosen for the predicted sequence. Amino acid residues that aligned through all three organisms were left unchanged in the predicted sequence. (**B**) The experimental MS/MS spectrum from the LC/MS/MS analysis of a triply charged peptide from ostrich bone protein extract that matched to the predicted sequence GPAGP (OH)PGKNGDDGEAGKP(OH)GRP(OH)GER and contained three hydroxyproline residues. (**C**) The MS/MS spectrum of the synthetically derived triply charged peptide of the same sequence for confirmation. The typical b- and y- fragment ions from the fragmentation pattern of the experimental peptide align very well with the synthetic peptide, validating the sequence interpretation.

from multiple related extant taxa. This approach produced a manageable number of theoretical protein sequences. The predicted peptide fragmentation pattern from these theoretical protein sequences were then compared with the fragmentation patterns of additional peptides derived from extracts of fossilized bone that did not match peptides in public sequence databases (fig. S1).

As a proof of concept for this approach, we investigated collagen sequences from femur bone extracts of an ostrich (*Struthio camelus*), an extant organism whose genomic sequence has not yet been evaluated and whose protein sequences are not available in protein databases. Collagens are the most abundant proteins in bone (>90%) and have specific posttranslational modifications (*12, 13*), and their longevity has already been demonstrated in fossils (*3, 14*). Collagen proteins are also highly conserved. For example, the sequence identity for collagen α1 type 1 (α1t1) from human (*Homo sapiens*) to frog (*Xenopus laevis*) is 81%, and the sequence identity between human and bovine (*Bos taurus*) is 97%, an extraordinarily high similarity. Using the Sequest algorithm (*15*), we identified 87 tryptic peptide spectra by microcapillary LC/MS/MS (LC, liquid chromatography) representing peptide sequence matches to extant related organisms in protein databases, primarily collagen α1t1 from chicken (*Gallus gallus*) (table S1).

Table 1. Extinct mastodon collagen proteins identified by LC/MS/MS, showing identity to extant organisms. A list of collagen proteins from 160,000- to 600,000-year-old mastodon fossilized bone is shown, including the number of peptide spectra identified, the amino acid coverage, and the organism identity.

Protein name	Organism identity	Number of peptide spectra	Amino acid coverage
Collagen α1(I)-chain precursor	Dog, bovine, human, chimp	24	20%
Similar to α2t1 collagen	Dog, human	15	10%
Similar to α2t1 collagen	Elephant	12	9%
Similar to collagen α1(IV)-chain precursor	Bovine	3	4%
α1t1 collagen	Human	2	2%
α1t2 collagen isoform 1	Human	3	4%
Collagen α2(I) chain	Human	4	6%
Similar to collagen α1t1	Elephant	2	3%
Collagen α1(I) chain	Mouse	2	2%
α1t2 collagen	Newt	2	5%
Similar to α2t1 collagen	Chicken	3	4%
Similar to collagen α1(I)-chain precursor	Chimp	2	2%
α1t1 collagen	Newt	2	3%

A Peptide Sequences Unique to Mastodon	Protein	Xcorr	Sp
GSEGPQGTR	collagen α1t1	2.22	1187
GAPGPQGP*G*GAP*GPK	collagen α1t1	3.44	1017
EGAPGSEGAPG*RDGAIGPK	collagen α1t1	2.86	580
GLTGPIGPP*GPAGAP*GDKGEG*GPSGPAGPTGAR	collagen α1t1	5.63	715
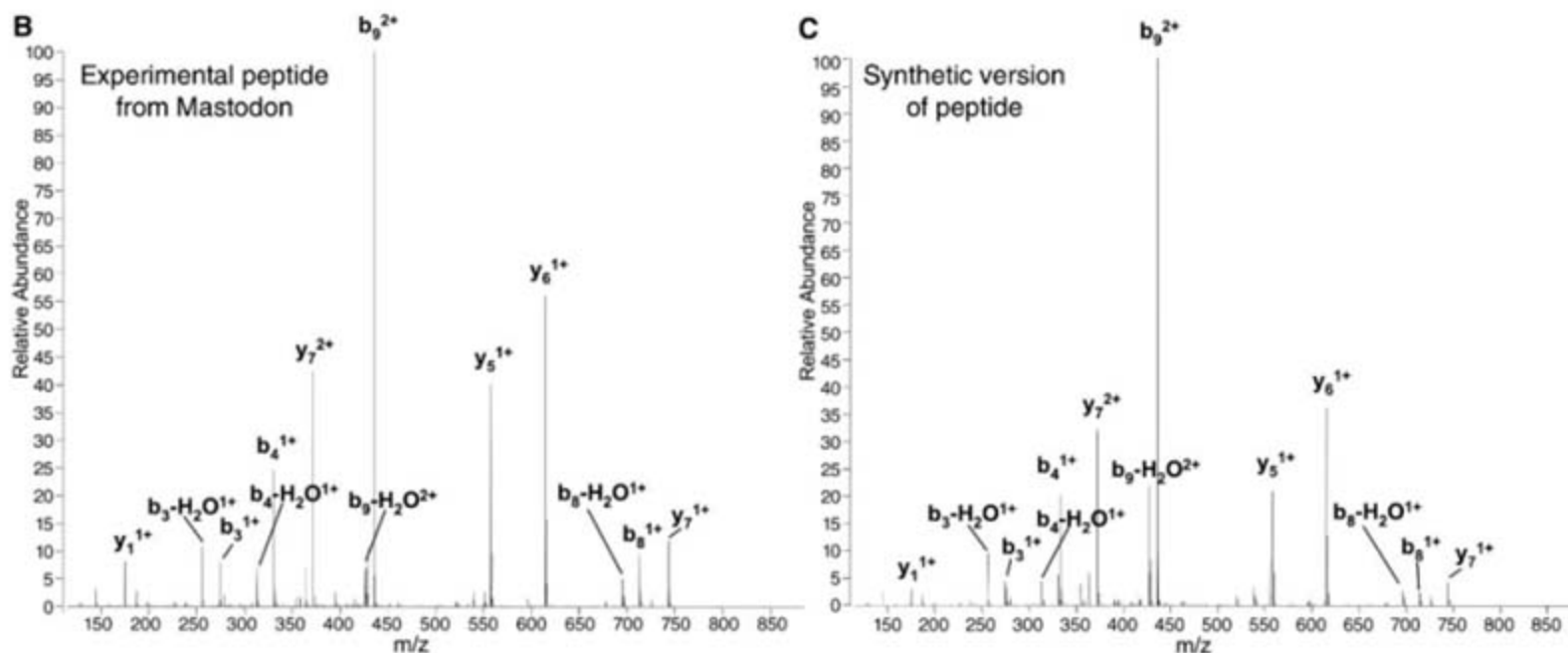
Fig. 2. Collagen peptide sequences unique to extinct mastodon identified by LC/MS/MS. (**A**) The four collagen α1t1 peptide sequences found by the approach that are unique to ancient mastodon. Xcorr (cross-correlation score) and Sp (preliminary score) represent the scores resulting from database searching against protein databases using Sequest. The asterisk represents the hydroxylation site after the posttranslationally modified residue. (**B**) An example of the experimental MS/MS spectrum of a doubly charged tryptic peptide for the collagen α1t1 peptide sequence GSEGPQGTR from the LC/MS/MS analysis of mastodon fossilized bone extract identified from a Sequest search against a theoretical collagen protein database. (**C**) The synthetic version of the same peptide sequence. All major ions from the experimental spectrum align very well with the ions from the synthetic version, validating the sequence.

In addition to obtaining sequence data, a further benefit of using MS is that posttranslational modifications (16) of the proteins can be determined. Approximately 50% of proline residues, 15% of lysines, and 10% of glycines in the collagen peptides were hydroxylated. Hydroxyproline stabilizes the triple helical confirmation of collagen in fibrillar structures (12), and hydroxylysine cross-links individual collagen molecules (13), although the function of glycine hydroxylation has not been reported. In some cases, the technology used here could not determine posttranslational modifications resulting in very small mass shifts nor could it distinguish isobaric amino acid residues. Approximately 33% of the sequence for collagen α1t1 and 16% for collagen α2t1 were

identified from ostrich bone extracts through identical matches to collagen in taxa whose sequences are present in protein databases. Many experimental factors can result in a failure to obtain complete protein coverage, including proteolysis, chromatography, and ionization efficiency; however, some peptide sequences could have been missed because of the evolutionary divergence of amino acids from sequences of taxa in current protein databases.

To address this latter possibility, we generated aligned sequences obtained from chicken collagen α1t1 (the most closely related sequence to that of ostrich in the public database) with those from frog and Japanese newt (Cynops pyrrhogaster) (the next most closely related sequences in the database). For predicted tryptic

fragments where one or more of these three taxa diverged at more than one residue, we generated a set of theoretical peptide/protein sequences that included the exact sequence in regions where all three species were identical and various combinations of the observed variant amino acids at residues where the three species diverged (Fig. 1A). We assumed that differences between chicken and ostrich are most likely to occur at residues that have been observed to drift from chicken to frog and newt. Because chicken is phylogenetically closer to ostrich than frog or newt, we chose the residue observed in chicken as the most likely residue when all three species differed at a location but chose the majority residue where two out of three were identical at a given position. For this example, we predicted a theoretical drift from chicken that was likely to be observed in a related species such as ostrich. This sequence (as well as other sequences predicted in an analogous manner) matched MS/MS fragmentation patterns from ostrich bone extract as well as a synthetic peptide created for sequence validation (Fig. 1, B and C).

Additional theoretical sequences were generated for misaligned residues by using point-assisted mutation (PAM) matrices that predict changes in amino acid residues through evolution (17), rather than choosing a residue from one of the three initial organisms (18). In addition to sequence matches to related organisms, the approach found six additional collagen peptide sequences that were unique to extant ostrich

Table 2. 68-million-year-old *T. rex* collagen peptide sequences identified by LC/MS/MS. Organism identity indicates the extant organisms to which the MS/MS fragmentation pattern perfectly aligned. Xcorr, Sequest cross-correlation score; Sp, Sequest preliminary score; *, hydroxylation site after a modified residue. The majority of collagen sequence matches from *T. rex* align uniquely with chicken from publicly available protein databases.

Peptide sequence	Protein	Organism identity	Xcorr	Sp
GATGAP*GIAGAPG*FP*GAR	Collagen α1t1	Chicken, frog	3.77	1099
G*AAGPP*GATGFP*GAAGR	Collagen α1t1	Newt, fish, mouse	3.74	797
GVQGPP*GPQGPR	Collagen α1t1	Chicken	2.54	865
GLPGESGAVGPAGPIGSR	Collagen α2t1	Chicken	2.99	479
GVVGLP*GQR	Collagen α1t1	Multiple organisms	2.55	500
GLVGAPGLRGLPGK	Collagen α1t2	Frog	2.28	410
GAPGPQG*PAGAP*GPK	Collagen α1t1	Newt	2.14	272
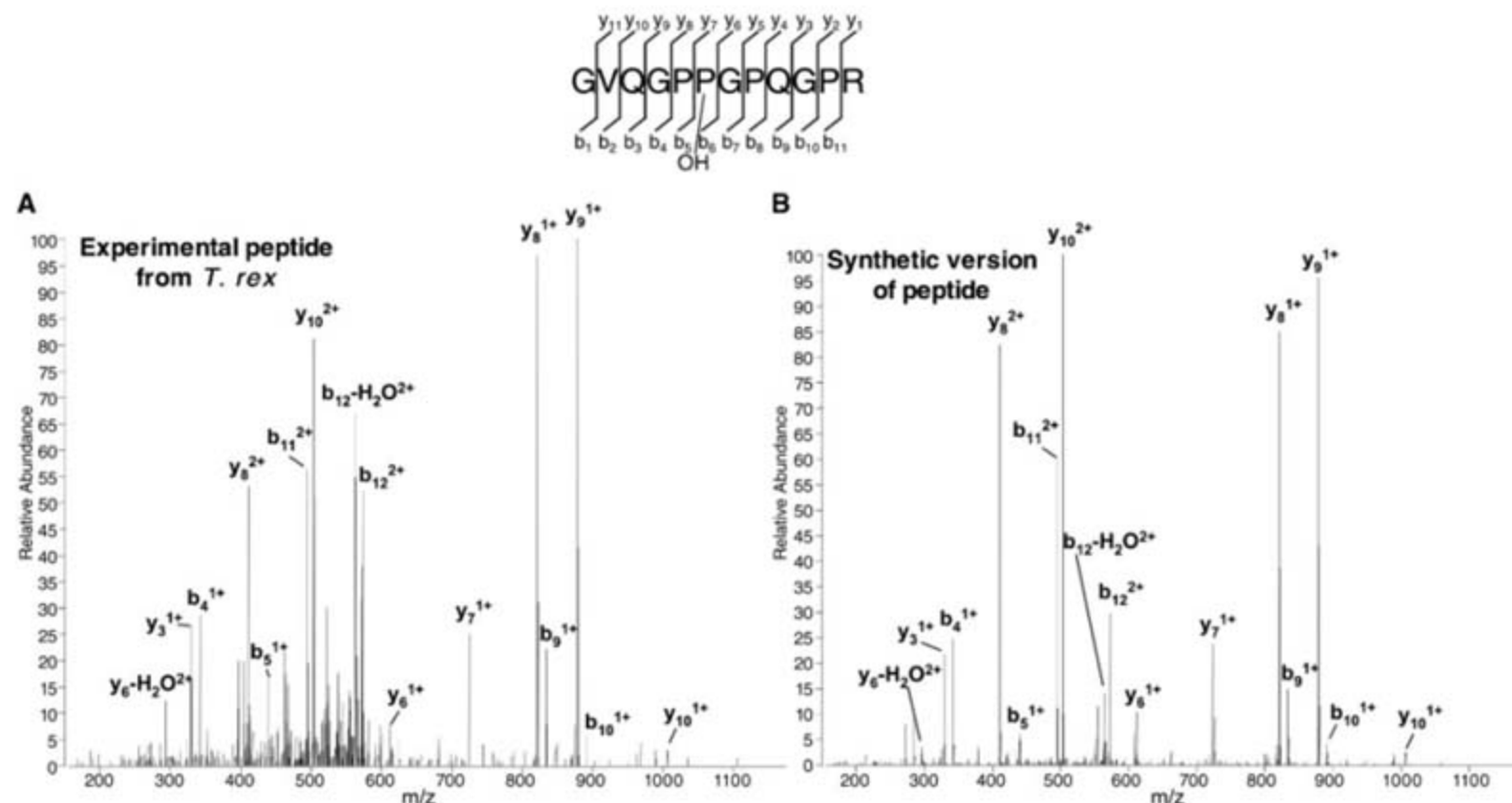
Fig. 3. The LC/MS/MS fragmentation pattern from a 68-million-year-old *T. rex* peptide. (**A**) The experimental MS/MS spectrum for the *T. rex* doubly charged hydroxylated tryptic peptide sequence GVQPP(OH)GPQGPR from femur bone extract identified by LC/MS/MS. (**B**) The synthetic version of the same sequence. All major fragment ions from the experimental spectrum are in very good alignment with ions from the synthetic version, confirming the sequence. This molecular sequencing evidence of protein from a 68-million-year-old fossilized bone demonstrates excellent preservation of the *T. rex* femur and the high sensitivity of state-of-the-art MS technology.

and had been missed in the comparison with public databases (table S2). Of these six sequences, four were determined using the organism weighted-consensus method and two were determined using the PAM weighted-consensus method. Although this approach only increased the total sequence coverage of ostrich collagen from ~33% to ~39%, it revealed sequences in ostrich collagen that differed from those in chicken or in other species in the database, thus providing a means to assess evolutionary divergences between these species.

For mastodon sequences, we sampled extracts from mastodon long-bone fragments dating to 160,000 to 600,000 years ago that preserved soft tissues (9, 19). 74 tryptic peptide MS/MS spectra matched collagen sequences from extant mammalian organisms in the protein database (Table 1). The sequence matches resulted in approximately 32% coverage for collagen α1t1 and 29% for collagen α2t1. The fraction of glycine with hydroxylation in the mastodon collagen peptides was somewhat higher (~50%) than that observed in ostrich collagen, whereas the fraction of hydroxylated proline and lysine residues was similar to that in ostrich. It is possible that the subset of peptides with hydroxylation on glycine is resistant to proteolysis, thus explaining the enrichment. Alternatively, enzymatic glycine hydroxylation may be more active in the mastodon than in the ostrich, or the oxidation may have occurred nonenzymatically over hundreds of thousands of years. The mastodon sequences obtained were more closely related to collagen sequences from dog, bovine, human, and elephant than to nonmammalian taxa, as expected.

We used the same approach as for ostrich to identify collagen peptide sequences unique to mastodon. We compared drifts in sequences between vertebrate mammals, including human, dog (*Canis familiaris*), and mouse (*Mus musculus*) as well as PAM to generate additional theoretical collagen sequences centered on extant mammalian species and added them to our protein database containing theoretical ostrich sequences. Elephant (*Loxodonta africana*), the most closely related extant taxon (20), was not used to generate predicted collagen sequences because incomplete peptide sequence fragments translated from genomic data are available in public protein databases. Other combinations of mammalian organisms such as bovine were also used to generate additional predictions. Four additional mastodon peptide sequences from fragmentation patterns matched the theoretical sequences; two from the organism weighted-consensus method and two from the PAM weighted-consensus method (Fig. 2A). Those sequences were confirmed by comparing their fragmentation patterns to those obtained with synthetic peptides (Fig. 2, B and C).

The additional sequences increased the amino acid coverage of mastodon collagen α1t1 from approximately 32% to 37%, a comparable increase to that for the ostrich collagen. We thus obtained nearly as much coverage of collagen

sequence from the 160,000- to 600,000-year-old fossilized bone (37%) as from a freshly collected ostrich bone (39%).

For *T. rex* sequences, we analyzed proteins from the femur of a 68-million-year-old *T. rex* (MOR 1125) recovered from the base of the Hell Creek Formation. This bone preserves soft tissues (10). Previous attempts to obtain protein or DNA sequences from such ancient fossils have failed because of extremely low concentrations of organic material remaining after extraction and because of degradation or modification of the remaining organic materials (7). Protein extracts from the *T. rex* were prepared as for ostrich and mastodon; however, the tryptic peptides required multiple purification steps (solid-phase extraction, strong cation exchange, and reversed-phase microchromatography) in order to eliminate a rust-colored coextracting contaminant and to increase the concentration of peptidic material. Three sequences from initial LC/MS/MS experiments from the *T. rex* samples indicated the presence of the iron-containing metalloenzyme nitrile hydratase beta, derived from soil bacteria *Rhodococcus sp.* and involved in biodegradation (21). Two peptides from a collagen adhesion protein from *Solibacter usitatis* were also sequenced. Microbial contamination was not significant, most likely because of the deep burial of the fossilized bones in the strata of the Hell Creek Formation.

The MS/MS spectra obtained from processed *T. rex* bone extracts revealed seven total collagen peptide sequences that could be aligned with predicted fragmentation patterns of collagen α1t1, α2t1, or α1t2 sequences from extant vertebrate taxa in the public protein database (Table 2). These sequences could be reproduced from multiple LC/MS/MS experiments; however, different peptides were sequenced from five different sample preparations of *T. rex* protein extract over a 1.5-year period. The last two extractions yielded less sequence information than earlier extractions, probably because of degradation of the fossil over time after removal from its well-preserved native environment (22). As in the extant ostrich and extinct mastodon, most of the peptides contained hydroxyl modifications on proline, lysine, or glycine residues. Sediment and buffer control samples were analyzed, and no sequences from collagen were found, although bacterial peptides were also present in sediment.

A BLAST alignment and similarity search (23) of the five *T. rex* peptides from collagen α1t1 as a group against the all-taxa protein database showed 58% sequence identity to chicken, followed by frog (51% identity) and newt (51% identity). The small group of peptide sequence data reported here support phylogenetic hypotheses suggesting that *T. rex* is most closely related to birds among living organisms whose collagen sequence is present in protein databases (24–26). The collagen sequences from other closely related extant taxa such as alligator (*Alligator sinensis*) and crocodile (*Crocodylus acutus*) are not present

in current protein databases. If all sequences were consistent with a single extant organism, it might indicate that the samples or our experiments were contaminated. However, we identified regions of sequence that align uniquely with multiple related vertebrate taxa in protein databases. The highly conserved nature of collagen proteins results in very limited regions that do not overlap, and the sequence alignments vary by only one or two amino acids, even in distantly related organisms. Because these peptides are all derived from the same bone matrix, one would have to make the argument for multiple contamination events from organisms, such as newt, that are not native to Hell Creek environments and have never been inside the buildings that have housed these bone samples. For further validation of the sequence data, Fig. 3 shows one of the experimental *T. rex* sequences, GVQGPP(OH)GPQGPR (27), that matched to chicken collagen α1t1 and the synthetic version of the peptide. The experimental spectrum shows lower signal intensity and more chemical noise than the synthetic peptide, which is not surprising because the spectra were derived from 68-million-year-old endogenous proteins. The signal intensities of the mass spectra indicate that only low or subfemtomole levels of peptides were produced from tryptic digestions of approximately 30 mg of bone protein extract. Peptide sequences unique to *T. rex* were not found, most likely because few peptides were available for sequencing as compared to the ostrich and mastodon samples. In support of these results and data shown here and by Schweitzer *et al.* (11), in situ localization with avian antibodies to collagen type 1 shows the presence of collagen, which disappears after treatment with collagenase (11).

The ability to sequence intact peptides from a 68-million-year-old source is attributed to several factors, including the exceptional preservation of the soft tissues from the Hell Creek environment, the fresh preparation of the fossil samples without curation or preservation (22), and the advancements in the sensitivity of MS technology over the past decade. The fact that sequenceable collagen was very abundant in the mastodon sample, which could be approximately half a million years old, also sheds light on the fact that sequenceable protein lasts much longer than 1 million years.

As technologies become more refined and protein extraction techniques are optimized, more informative material may be recovered. This holds promise for future work on other fossil material showing similar preservation, but also demonstrates a method for obtaining protein sequences from rare or endangered extant organisms whose genomes have note been sequenced. The MS- and bioinformatics-based approach we have used can be applied not only to obtain sequences from extinct organisms, but also to obtain protein sequences from extant organisms whose genomes have not been sequenced and to discover mutations in diseased tissues such as cancers.

References and Notes

1. M. Mann, O. N. Jensen, *Nat. Biotechnol.* **21**, 255 (2003).
2. B. Domon, R. Aebersold, *Science* **312**, 212 (2006).
3. M. H. Schweitzer *et al.*, *J. Mol. Evol.* **55**, 696 (2002).
4. C. M. Nielsen-Marsh *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4409 (2005).
5. C. M. Nielsen-Marsh *et al.*, *Geology* **30**, 1099 (2002).
6. P. H. Ostrom *et al.*, *Geochim. Cosmochim. Acta* **70**, 2034 (2006).
7. T. Lindahl, *Nature* **362**, 709 (1993).
8. M. J. Collins *et al.*, *Geology* **28**, 1139 (2000).
9. C. L. Hill, *Quat. Int.* **142-143**, 87 (2006).
10. M. H. Schweitzer, J. L. Wittmeyer, J. R. Horner, J. K. Toporski, *Science* **307**, 1952 (2005).
11. M. H. Schweitzer *et al.*, *Science* **316**, 277 (2007).
12. A. Bhattacharjee, M. Bansal, *Int. U. Biochem. Mol. Biol. Life* **57**, 161 (2005).
13. L. Knott, A. J. Bailey, *Bone* **22**, 181 (1998).
14. N. Tuross, M. L. Fogel, P. E. Hare, *Geochim. Cosmochim. Acta* **52**, 929 (1988).
15. J. K. Eng, A. L. McCormack, J. R. Yates III, *J. Am. Soc. Mass Spectrom.* **5**, 976 (1994).
16. M. Mann, O. N. Jensen, *Nat. Biotechnol.* **21**, 255 (2003).
17. M. O. Dayhoff *et al.*, *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3 (National Biomedical Research Foundation, Washington, DC, 1978).
18. A complete description of the predicted sequence generation is available as supporting material on *Science* Online.
19. M. H. Schweitzer, J. L. Wittmeyer, J. R. Horner, *Proc. R. Soc. London Ser. B* **274**, 183 (2007).
20. H. Yang, E. M. Golenberg, J. Shoshani, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1190 (1996).
21. M. M. Tauber *et al.*, *Appl. Environ. Microbiol.* **66**, 1634 (2000).
22. M. Pruvost *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 739 (2007).
23. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **215**, 403 (1990).
24. P. M. O'Connor, L. P. Claessens, *Nature* **436**, 253 (2005).
25. F. V. Mariani, G. R. Martin, *Nature* **423**, 319 (2003).
26. M. H. Schweitzer, J. L. Wittmeyer, J. R. Horner, *Science* **308**, 1456 (2005).
27. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
28. We thank J. Wittmeyer, J. Luo, R. Mecham, M. Berne, M. Vander Heiden, H. Christofk, and B. Zheng for experimental help and helpful discussions; W. Lane and S. Nagpal for software help; J. R. Horner for providing *T. rex* and mastodon fossils; and C. Hill for taphonomic data for mastodon. We also thank NSF (awards EAR-0634136 and EAR-0541744), the Paul F. Glenn Foundation, and the David and Lucile Packard Foundation for providing funds.

Lymphotoxin β Receptor–Dependent Control of Lipid Homeostasis

James C. Lo,[1]* Yugang Wang,[2]* Alexei V. Tumanov,[2]* Michelle Bamji,[3] Zemin Yao,[3] Catherine A. Reardon,[2] Godfrey S. Getz,[2]† Yang-Xin Fu[1,2]†

Hyperlipidemia, one of the most important risk factors for coronary heart disease, is often associated with inflammation. We identified lymphotoxin (LT) and LIGHT, tumor necrosis factor cytokine family members that are primarily expressed on lymphocytes, as critical regulators of key enzymes that control lipid metabolism. Dysregulation of LIGHT expression on T cells resulted in hypertriglyceridemia and hypercholesterolemia. In low-density lipoprotein receptor–deficient mice, which lack the ability to control lipid levels in the blood, inhibition of LT and LIGHT signaling with a soluble lymphotoxin β receptor decoy protein attenuated the dyslipidemia. These results suggest that the immune system directly influences lipid metabolism and that LT modulating agents may represent a novel therapeutic route for the treatment of dyslipidemia.

Atherosclerosis results from a combination of dyslipidemia (dysregulation of lipid levels in the blood) and inflammation-mediated pathology of the vasculature (*1–3*). In particular, systemic inflammation is often associated with hyperlipidemia, although the exact mechanistic link is unclear (*4*). Members of the tumor necrosis factor superfamily (TNFSF) of cytokines are involved in a wide array of biological processes, including immune function, cell death and survival, and atherosclerosis. Among these, lymphotoxin (LT) and LIGHT are potent proinflammatory ligands expressed primarily on T cells, and both bind to lymphotoxin β receptor (LTβR), a tumor necrosis factor receptor superfamily member. Disruption of LTβR signaling in mice results in a wide array of phenotypes including lymph node aplasia,

impaired splenic microarchitecture, autoimmunity, viral infection, and grossly impaired immunoglobulin (Ig) responses (*5–7*).

Recent clinical, genetic, and epidemiological studies have linked LT to the development of atherosclerosis and myocardial infarction (*8, 9*). However, the contribution of LT to lipid homeostasis—a likely influencing factor in atherogenesis—is not established. In one inflammatory mouse model, increased expression of LIGHT on T cell lineage in *Lck* LIGHT transgenic (Tg) mice culminates in the development of systemic inflammation, hepatomegaly, and inflammatory bowel disease (*10, 11*). Lymphocytes can have intimate contact with hepatocytes, which raises the possibility that some T cell–derived ligands such as LT and LIGHT may deliver signals that regulate various liver functions. Indeed, LT and LIGHT have been shown to impart signals necessary for liver regeneration as well as for pathogenesis of hepatitis (*12, 13*).

Given the observed effects of LT and LIGHT on the liver, and given that lipid homeostasis

[1]Committee on Immunology, University of Chicago, Chicago, IL 60637, USA. [2]Department of Pathology, University of Chicago, Chicago, IL 60637, USA. [3]Department of Biochemistry, Microbiology and Immunology, University of Ottawa Heart Institute, Ottawa, Ontario K1H 8M5, Canada.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: yfu@uchicago.edu (Y.-X.F.); getz@bsd.uchicago.edu (G.S.G.)

Fig. 1. T lymphocyte–derived LIGHT promotes dyslipidemia. (**A** to **D**) Wild-type (WT) and LIGHT Tg mice were fed a normal chow diet (**A** and **B**) or Western-type diet (WTD) (**C** and **D**) for 2 weeks and then assayed for plasma total cholesterol and triglycerides. (**E**) WT and LIGHT Tg spleen and lymph node cells (10⁷ per mouse) were transferred into sublethally irradiated (600 rad) WT mice that were then fed WTD; after 4 weeks, plasma total cholesterol levels were determined. (**F**) LIGHT Tg mice on a normal chow diet were injected with control human Ig (hIg) or LTβR-Ig on day 0; after 1 week, plasma total cholesterol was determined. Each individual symbol represents values from WT (triangles) and LIGHT Tg (circles) mice. Means are plotted as columns. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

is tightly regulated by the liver, we tested whether lipid metabolism might be perturbed in the *Lck* LIGHT Tg mice. At 4 to 8 weeks of age, LIGHT Tg mice displayed hypercholesterolemia ($P < 0.001$) even on a normal diet (Fig. 1A) and triglyceride levels were also elevated in LIGHT Tg mice relative to wild-type mice ($P = 0.02$) (Fig. 1B). This hyperlipidemic phenotype was further exaggerated when the mice were placed on a high-fat, high-cholesterol diet, with LIGHT Tg mice developing hypertriglyceridemia ($P < 0.01$) and worsened hypercholesterolemia ($P < 0.01$) (Fig. 1, C and D). In contrast, the plasma lipid levels of wild-type mice did not change on the same diet (Fig. 1, A to D).

Although liver LIGHT Tg CD4 and CD8 T cell numbers (fig. S1A) as well as LIGHT expression (fig. S2) were increased in 4- to 8-week-old *lck* LIGHT Tg mice, inflammatory cell infiltration and liver cell damage were not obvious at this point (fig. S1, B and C). Notably, these increases occurred earlier than the development of inflammatory bowel disease in these animals (fig. S1F). This raised the possibility that LIGHT itself, or activated T cells, might facilitate hyperlipidemia. Using transfer of purified LIGHT Tg T cells to wild-type mice to test this hypothesis, we observed increased plasma cholesterol levels in the recipient animals (Fig. 1E and fig. S3), which could be corrected by inhibition of LIGHT signaling with the use of LTβR-Ig fusion protein (Fig. 1F). These results revealed that LIGHT expressed on T lymphocytes was capable of regulating lipid homeostasis.

To further address the potential mechanism of LIGHT-mediated dyslipidemia, we crossed LIGHT Tg mice with mice lacking each of the two defined receptors for LIGHT: the herpesvirus entry mediator (HVEM) and LTβR (6, 7). LIGHT-mediated dyslipidemia was still present and even worsened in LIGHT Tg/$HVEM^{-/-}$ mice but was largely corrected in LIGHT Tg/$LT\beta R^{-/-}$ mice (Fig. 2). These data suggest that the LIGHT-LTβR interaction, but not the LIGHT-HVEM interaction, was essential for the dyslipidemia in LIGHT Tg mice. In contrast, HVEM signaling may have had a modest protective influence on dyslipidemia in most mice.

Two likely target organs known to regulate lipid homeostasis are the liver and intestine. We focused our initial attention on the liver as the most likely organ to influence plasma lipid balance, because there was no evidence of impaired lipid absorption in the gut of these animals (fig. S1E). Hepatic gene expression profiles of wild-type and LIGHT Tg mice were first generated by Affymetrix microarray analysis (14). Of more than 45,000 hepatic genes, 385 were identified with expression levels that were altered by a factor of 5 or more. These genes were subdivided into specific pathways, with particular attention to those known to regulate lipid and cholesterol metabolism. In particular, relative expression of hepatic lipase (HL) was decreased by a factor of 20 in the livers of LIGHT Tg mice (Fig. 3, A and

B, and fig. S4). HL is a key enzyme in lipid metabolism by the liver, hydrolyzing triglycerides and phospholipids associated with plasma lipoproteins (15). HL may function both as a lipolytic enzyme and as a ligand for hepatic receptor-based uptake of triglyceride-rich lipoproteins (15). Inactivation of the *HL* gene in low-density lipoprotein receptor (LDLR)–deficient mice has

Fig. 2. LIGHT-dependent dyslipidemia is mediated by LTβR signaling. (**A** to **D**) Mice of the indicated genotypes were fed WTD for 2 weeks and were then assayed for plasma total cholesterol [(A) and (C)] and triglycerides [(B) and (D)]. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.
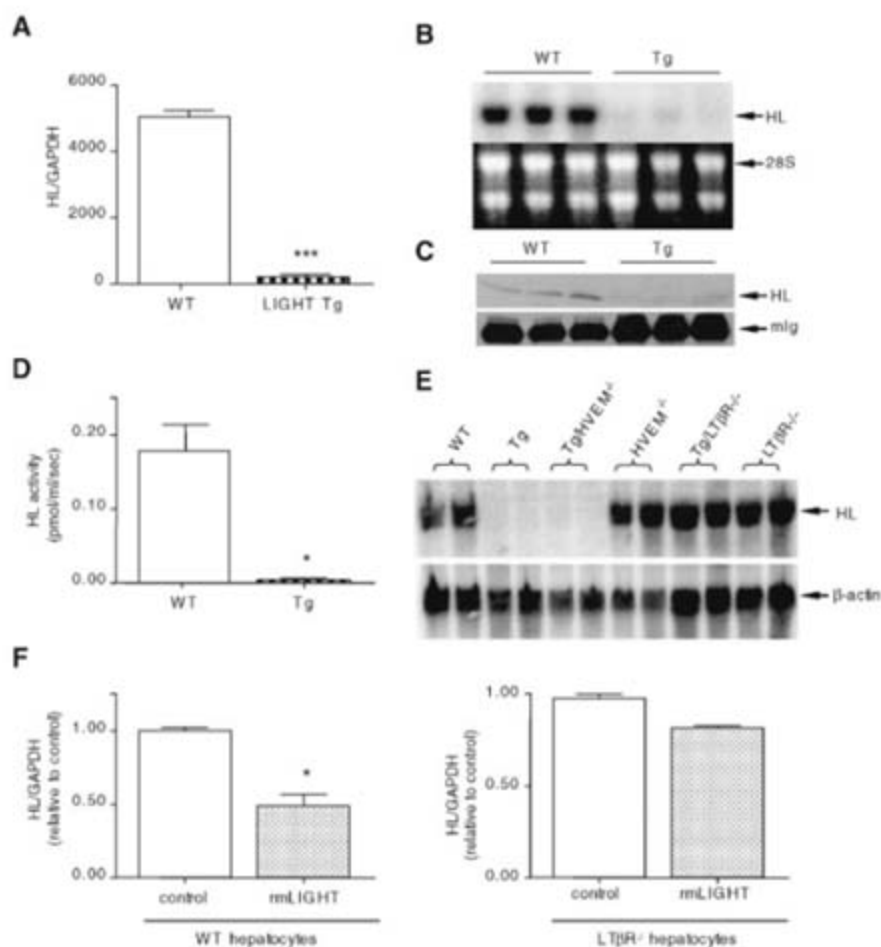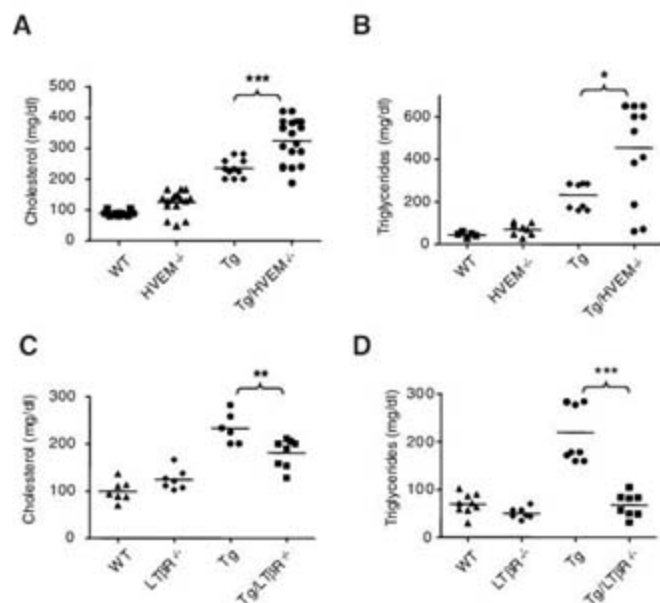
Fig. 3. Dysregulated LIGHT expression on T cells reprograms hepatic gene expression. (**A**) *HL* mRNA level expression by real-time PCR. ***$P < 0.001$. (**B**) Northern blot analysis of HL expression. (**C**) Analysis of HL protein expression level in post-heparin plasma by Western blot. (**D**) Post-heparin plasma HL enzyme activity analysis using the CONFLUOLIP hepatic lipase kit. *$P < 0.05$. (**E**) LIGHT-mediated repression of HL is dependent on the LTβR, not HVEM signaling. Shown is a representative Northern blot on the mice of the indicated genotypes. (**F**) LIGHT can directly repress HL expression in hepatocytes. Primary hepatocytes from WT (left) or $LT\beta R^{-/-}$ (right) mice were incubated with or without recombinant mouse LIGHT overnight, and their *HL* mRNA levels were quantified by real-time reverse transcription PCR. *$P < 0.05$; experiments were performed three times.

been found to result in elevation of very-low-density lipoprotein (VLDL) and dyslipidemia (*16*), whereas overexpression of HL acts to substantially decrease triglycerides and lipoprotein cholesterol, especially in the ApoB-containing lipoproteins (*17*). The change of *HL* gene expression in LIGHT Tg mice (Fig. 3B) correlated very well with the protein expression level and enzyme activity in the post-heparin plasma (Fig. 3, C and D). The levels of *HL* mRNA expression were increased in the livers of LIGHT Tg/$LT\beta R^{-/-}$ mice relative to LIGHT Tg mice (Fig. 3E). In contrast, the enzyme activity of HL and its mRNA expression in LIGHT Tg/$HVEM^{-/-}$ mice remained at levels seen in LIGHT Tg mice (Fig. 3E and fig. S5). Taken together, these results suggest that LIGHT-LTβR interactions inhibit the expression of HL in a way that could explain the hyperlipidemia in LIGHT Tg mice.

Taking into account the effective normalization of lipid levels in LIGHT Tg mice by LTβR-Ig treatment (Fig. 1F), we tested whether inhibition of LTβR signaling could restore cholesterol levels in hypercholesterolemic $HL^{-/-}$ mice. The cholesterol levels were unchanged after weekly treatment of HL-deficient mice with LTβR-Ig reagent for 8 weeks (fig. S6), although

a rapid change was detected in LIGHT Tg mice by the same treatment (Fig. 1F). These data provide further support for the importance of LTβR signaling in regulation of HL expression and consequent dyslipidemia.

Thus far, our data support the hypothesis that LIGHT could potentially signal via the LTβR on hepatocytes to directly inhibit the expression of HL. To directly assess this possibility, we stimulated primary wild-type and $LT\beta R^{-/-}$ hepatocytes with recombinant mouse LIGHT; the result was a significant depression of the expression of HL in primary wild-type hepatocytes ($P < 0.05$) (Fig. 3F) but not in the primary hepatocytes lacking LTβR expression. To test whether direct contact between hepatocytes and T cells was responsible for control of HL expression, we incubated hepatocytes and LIGHT-expressing T cells together or separated with a transwell culture. In these experiments, inhibition of HL expression was evident only when hepatocytes and LIGHT Tg T cells could be in direct contact (fig. S7B), which further suggested that LIGHT signaling via direct interaction with LTβR is required for its effects on HL expression in hepatocytes.

To this point, our data suggest a pathogenic role for LIGHT-induced LTβR signaling in

plasma lipid homeostasis. To test this in a separate model, we analyzed $LDLR^{-/-}$ mice, which also display dyslipidemia (*18*). Hyperlipidemia is less severe when $LDLR^{-/-}$ mice are crossed with $RAG1^{-/-}$ mice (*19*), so it is possible that the immune system might participate in lipid regulation. Potentially this might be initiated by dyslipidemia induced by the LDLR deficiency; in such a scenario, the immune system would be activated and in turn would activate LTβR signaling and further promote hyperlipidemia by a positive feedback loop. To test this hypothesis, we first tested whether overexpression of LIGHT by T cells in $LDLR^{-/-}$ mice could promote hyperlipidemia. Transfer of bone marrow cells from LIGHT Tg mice to $LDLR^{-/-}$ recipient mice enhanced hyperlipidemia relative to transfer of wild-type bone marrow (fig. S8). In contrast, the absence of LIGHT or LT ligand on transferred bone marrow cells ameliorated the hyper-cholesterolemia in $LDLR^{-/-}$ recipient mice, most prominently in the absence of LT ligand (fig. S9A). Interestingly, LDLR-deficient mice reconstituted with bone marrow from $LT^{-/-}$ donors, but not $LIGHT^{-/-}$ donors, expressed higher *HL* mRNA levels and HL enzyme activity relative to mice reconstituted with bone marrow from wild-type donors (fig. S9, B and C). These results imply that LTβR signaling is an important mediator of the pathogenesis of dyslipidemia in LDLR-deficient mice and that LT may be a more potent ligand in this model.

Although the numbers of activated (CD44$^+$) CD4$^+$ T cells did not show measurable change in $LDLR^{-/-}$ LT bone marrow transfer recipients (fig. S9E), the overall numbers of CD4 and CD8 T cells were reduced in these animals (fig. S9D), corresponding with reduced plasma cholesterol levels (fig. S9A). However, these bone marrow recipients did not display altered expression of macrophage ApoE (fig. S9F), which indicates that LTβR signaling might not work through the regulation of ApoE expression on macrophages.

Because our results showed a direct effect of LTβR signaling in dyslipidemia in $LDLR^{-/-}$ mice, we further tested the effects of therapeutic injection of LTβR-Ig protein. After weekly doses of LTβR-Ig for 12 weeks on a high-fat diet, $LDLR^{-/-}$ mice showed a substantial reduction in total cholesterol ($P < 0.001$) and triglycerides ($P < 0.01$) (Fig. 4, A and B). Fast protein liquid chromatography (FPLC) analyses of the plasma lipoproteins revealed that LTβR-Ig treatment reduced levels of VLDLc ($P < 0.01$) and LDLc ($P < 0.05$) without perturbing levels of HDLc (Fig. 4, C and D). However, in the absence of adaptive immune cells in $LDLR^{-/-}$ $RAG-1^{-/-}$ mice, or without natural killer (NK) T cells in $LDLR^{-/-}$ $CD1d^{-/-}$ mice, lipid levels are partially reduced and the same treatment was not effective in causing further reduction (fig. S10). These data imply that T cells, especially NK T cells, might be the major source of LT and/or LIGHT in promoting hyperlipidemia in $LDLR^{-/-}$ mice.
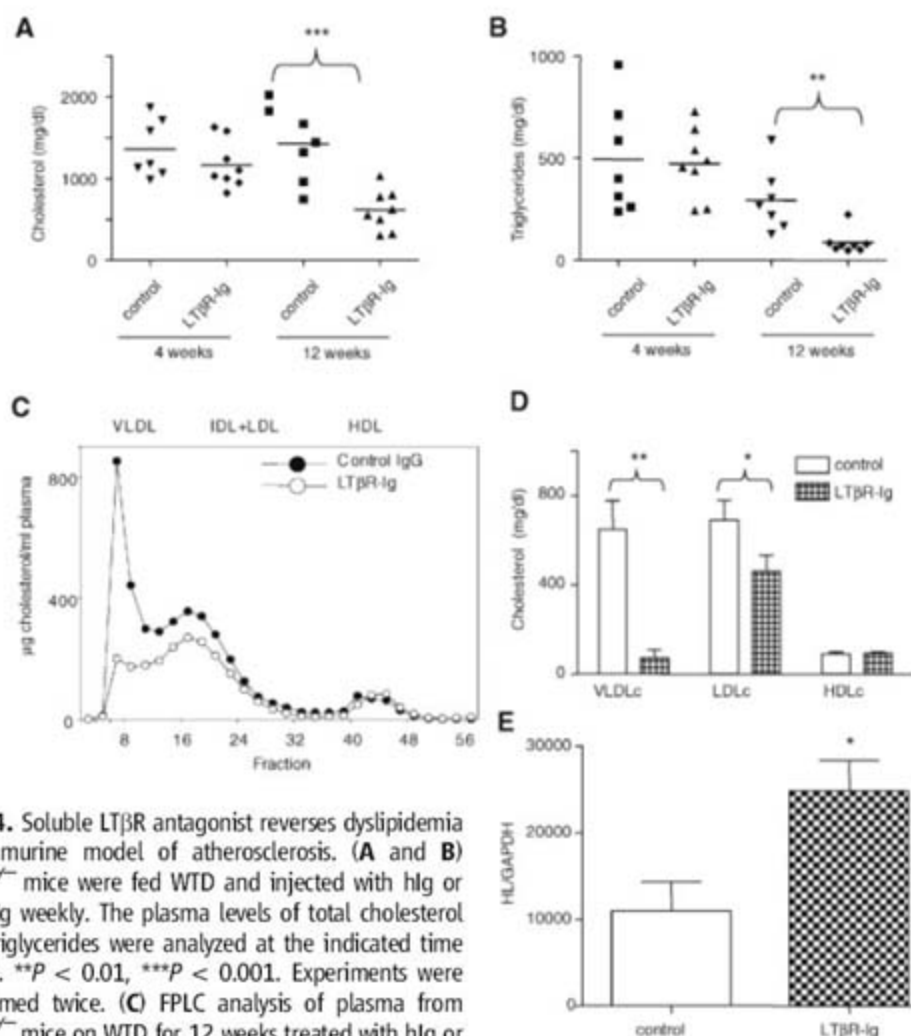
Fig. 4. Soluble LTβR antagonist reverses dyslipidemia in a murine model of atherosclerosis. (**A** and **B**) $LDLR^{-/-}$ mice were fed WTD and injected with hIg or LTβR-Ig weekly. The plasma levels of total cholesterol and triglycerides were analyzed at the indicated time points. **$P < 0.01$, ***$P < 0.001$. Experiments were performed twice. (**C**) FPLC analysis of plasma from $LDLR^{-/-}$ mice on WTD for 12 weeks treated with hIg or LTβR-Ig. (**D**) Lipoprotein cholesterol levels in $LDLR^{-/-}$ mice treated with hIg or LTβR-Ig. *$P < 0.05$, **$P < 0.01$. (**E**) HL expression levels were quantified by real-time PCR from livers of $LDLR^{-/-}$ mice on WTD treated with hIg or LTβR-Ig. *$P < 0.05$.

Real-time quantitative polymerase chain reaction (PCR) analyses revealed an increase of HL expression in the livers of mice treated with LTβR-Ig ($P < 0.05$) (Fig. 4E), which may have contributed to the decrease in VLDLc levels observed in these mice. These data suggest that antagonism of the LT-LIGHT signaling pathway can increase HL expression and ameliorate the dyslipidemia in LDLR-deficient mice.

This study points to a complex interaction between T cells and liver cells in regulating lipoprotein homeostasis, mediated by two TNFSF ligands (LT and LIGHT) and their receptors. Our data raise the prospect that T cells and T cell–derived ligands might contribute directly to the complex regulation of lipid homeostasis. However, genes other than *HL* are also likely to be involved in the LTβR-dependent effects we have observed and will clearly be the focus of future studies. Finally, it is possible that LT-modulating agents may represent a novel therapeutic approach for the treatment of dyslipidemia.

References and Notes

1. C. K. Glass, J. L. Witztum, *Cell* **104**, 503 (2001).
2. P. Libby, *Nature* **420**, 868 (2002).
3. G. K. Hansson, *N. Engl. J. Med.* **352**, 1685 (2005).
4. W. Khovidhunkit *et al.*, *J. Lipid Res.* **45**, 1169 (2004).
5. Y. X. Fu, D. D. Chaplin, *Annu. Rev. Immunol.* **17**, 399 (1999).
6. J. L. Gommerman, J. L. Browning, *Nat. Rev. Immunol.* **3**, 642 (2003).
7. C. F. Ware, *Annu. Rev. Immunol.* **23**, 787 (2005).
8. S. A. Schreyer, C. M. Vick, R. C. LeBoeuf, *J. Biol. Chem.* **277**, 12364 (2002).
9. K. Ozaki *et al.*, *Nature* **429**, 72 (2004).
10. J. Wang *et al.*, *J. Clin. Invest.* **108**, 1771 (2001).
11. R. B. Shaikh *et al.*, *J. Immunol.* **167**, 6330 (2001).
12. R. A. Anders, S. K. Subudhi, J. Wang, K. Pfeffer, Y. X. Fu, *J. Immunol.* **175**, 1295 (2005).
13. S. Anand *et al.*, *J. Clin. Invest.* **116**, 1045 (2006).
14. See supporting material on *Science* Online.
15. S. Santamarina-Fojo, H. Gonzalez-Navarro, L. Freeman, E. Wagner, Z. Nong, *Arterioscler. Thromb. Vasc. Biol.* **24**, 1750 (2004).
16. D. Barcat *et al.*, *Atherosclerosis* **188**, 347 (2006).
17. H. L. Dichek *et al.*, *J. Lipid Res.* **42**, 201 (2001).
18. C. A. Reardon, G. S. Getz, *Curr. Opin. Lipidol.* **12**, 167 (2001).
19. C. A. Reardon, L. Blachowicz, J. Lukens, M. Nissenbaum, G. S. Getz, *Arterioscler. Thromb. Vasc. Biol.* **23**, 1449 (2003).
20. We thank C. Lo for help with figures, L. Blachowicz and J. Lukens for technical assistance with mice, and L. Chen for reagents. Supported by NIH grants AI062026, CA097296, and DK58891 (Y.-X.F.) and HL 85516 (G.S.G.), Heart and Stroke Foundation of Ontario grant T5642 (Z.Y.), and American Heart Association grant 0730419Z (A.V.T.). J.C.L. is part of the Medical Scientist Training Program at the University of Chicago (grant 5 T32 GM07281). Array data were deposited into Gene Expression Omnibus (accession number GSE 6903).

Structural Basis for Substrate Delivery by Acyl Carrier Protein in the Yeast Fatty Acid Synthase

Marc Leibundgut,* Simon Jenni,* Christian Frick, Nenad Ban†

In the multifunctional fungal fatty acid synthase (FAS), the acyl carrier protein (ACP) domain shuttles reaction intermediates covalently attached to its prosthetic phosphopantetheine group between the different enzymatic centers of the reaction cycle. Here, we report the structure of the *Saccharomyces cerevisiae* FAS determined at 3.1 angstrom resolution with its ACP stalled at the active site of ketoacyl synthase. The ACP contacts the base of the reaction chamber through conserved, charge-complementary surfaces, which optimally position the ACP toward the catalytic cleft of ketoacyl synthase. The conformation of the prosthetic group suggests a switchblade mechanism for acyl chain delivery to the active site of the enzyme.

Acyl carrier protein (ACP) and related substrate shuttling domains are essential in many metabolic pathways, including fatty acid, polyketide, and nonribosomal protein biosynthesis (*1–3*). During fatty acid synthesis, substrates are linked via a thioester bond to the prosthetic phosphopantetheine (PPT) group of ACP. In the dissociated type II fatty acid synthase (FAS) systems of most bacteria and plants, ACP is a highly abundant, small protein, which sequentially transfers the acyl-chain intermediates between the different enzymes during the cyclic reaction (*1*). The type I FAS systems of mammals, fungi, and some bacteria comprise large, multifunctional enzymes into which ACP is integrated together with the catalytic domains (*4–6*). This leads to a drastically increased local concentration of ACP and of all catalytic domains and allows efficient catalysis by shuttling the reaction intermediates from one reaction center to the next (*7, 8*).

Previous structural studies of bacterial and plant type II ACPs and related carrier domains revealed that the proteins fold into a flexible all α-helical bundle (*9–15*). A hydrophobic cavity formed within the helix bundle of ACP shows high structural plasticity, allowing it to accommodate thioester-bound acyl-chains of different lengths (*12, 16*).

Limited structural information about interactions of ACP with catalytic enzymes and the mechanism of substrate delivery is currently available (*17*). Docking models, nuclear magnetic resonance and crystallographic experiments performed in the type II FAS system indicated the importance of the ACP recognition helix α2 for enzyme binding (*18–21*). In the structure of the fungal α6β6 2.6-megadalton FAS complex from *Thermomyces lanuginosus*, we observed weak electron density in both reaction chambers, possibly representing dynamically disordered ACP domains (*22*). In an attempt to gain additional information about substrate delivery by the ACP domain, we determined the structure of yeast FAS and visualized ACP stalled at the active site of ketoacyl synthase (KS), the substrate-condensing enzyme in the complex.

The *Saccharomyces cerevisiae* FAS crystal structure was solved by molecular replacement using the *T. lanuginosus* FAS coordinates (*23*). The comparison of the two homologous FAS structures reveals only minor structural differences (Fig. 1A, figs. S6 and S7). Both atomic models comprise the full set of catalytic domains involved in the fatty acid elongation cycle with an identical organization of the catalytic sites, which are pointing toward the inside of the reaction chamber. The small structural differences between the two FASs, which crystallize in different space groups and under distinct conditions, support the model for substrate shuttling without the need for large conformational changes of the enzymatic domains in the complex during the reaction cycle (*8, 22*). In yeast FAS, we were able to identify interpretable electron density for the ACP domains at the base of the chamber, close to the catalytic cleft of the KS domain (Fig. 1A and fig. S1). Each of the three ACPs per reaction chamber is N-terminally anchored to the chamber wall and C-terminally attached to the middle of the central wheel (Fig. 1B).

The structure of the yeast ACP domain is about twice the size (18 kD) of the free-standing bacterial ACP (Fig. 2A). Fungal ACP is all α-helical, and the four N-terminal helices superimpose well with structures of plant and bacterial homologs (*12, 16*), thereby defining the ACP core (Fig. 2A and fig. S2) (*16*). Serine $S^{180}_{ACP(Sc)}$ is located in a loop between helices α7 and α8, the latter corresponding to the ACP recognition helix (Fig. 2A). This residue has previously been identified as the attachment site of PPT by genetic and biochemical experiments (*24–27*).

Extending from the position of serine $S^{180}_{ACP(Sc)}$, unbiased electron density maps

Institute of Molecular Biology and Biophysics, ETH Zurich, 8092 Zurich, Switzerland.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: ban@mol.biol.ethz.ch

show the phosphate and pantoic acid moieties of the prosthetic group reaching into the active site cleft of KS (Fig. 2B). This conformation differs considerably from the one observed in the structures of *Escherichia coli* ACP in complex with acyl substrates (Fig. 2A) (*16*). In these complexes, the PPT moieties are rotated inward, and the acyl chains are buried in a deep, hydrophobic pocket formed by the ACP core, which may represent the transport form of acyl-ACP during substrate shuttling between the different catalytic domains. Notably, an empty hydrophobic pocket is recognizable also in the yeast ACP structure in complex with KS (Fig. 3A).

The N- and C-terminal tethering points on fungal ACP are opposite the site of the prosthetic group attachment, which maximizes the distance between the thiol group of the PPT arm, where the substrates are covalently bound, and the site where the flexible linkers are attached (Fig. 2A and 3A). This organization may also prevent the flexible linkers from interfering with ACP binding to the enzymes and substrate delivery.

The KS active site is located close to the approaching PPT arm of ACP and consists of a catalytic triad involving residues $C^{1305}_{KS(Sc)}$, $H^{1542}_{KS(Sc)}$, and $H^{1583}_{KS(Sc)}$ (Fig. 2B). The critical role of the catalytic cysteine has been established in yeast FAS by biochemical labeling and genetic experiments (*24*, *26*, *28*). Lysine $K^{1578}_{KS(Sc)}$ may also contribute to catalysis, because, in bacteria, a corresponding lysine is critical for decarboxylation (*29*).

We observe two major contact areas between fungal ACP and KS: The first involves the ACP core, including the ACP recognition helix α8, which contacts residues surrounding the entrance to the active site cleft of the KS core fold (Fig. 3B and figs. S2 and S3). This contact is consistent with the experiments, which demonstrate that the bacterial ACP recognition helix is important for enzyme binding (*18*–*21*). The second is formed between the additional part of the ACP domain and the large spoke of the central wheel, which consists of KS expansion segments (Fig. 3B). Notably, both subunits in the fungal KS dimer contribute to both ACP contact areas (figs. S3 and S4). The interactions of ACP with the condensing enzyme indicate a more directed mechanism for substrate delivery, rather than incidental binding of ACP-linked substrates to the KS active site.

Strong evidence indicates a physiological role for the observed interactions: (i) If we model the remaining β-alanine and cystamine moieties of the PPT arm in a sterically reasonable manner, starting from the observed pantoic acid moiety, the 18 Å–long prosthetic group extends directly into the KS catalytic cleft and terminates next to the catalytic cysteine $C^{1305}_{KS(Sc)}$ (Fig. 2B). (ii) ACP has patches of positively charged surface residues that interact with complementarily charged residues near the cleft entrance and in the large spoke of the central wheel (Fig. 3C). (iii) The residues that mediate the interactions are highly conserved in both the ACP domain and the reaction chamber (fig. S4D). (iv) We also observe weak electron density for ACP at a

similar position above the KS in *T. lanuginosus* FAS crystals (*22*).

Although tight and preferential binding of the ACP domain to one of the catalytic domains would interfere with efficient substrate shuttling, energetically balanced, low-affinity guiding interactions for all active sites would facilitate transfer of substrates and intermediates within the reaction chamber. It is not clear why the substrate-depleted multienzyme in the crystal has stalled at this particular stage of the reaction. One possibility is that under the crystallization conditions, which favor protein-protein interactions, ACP has a higher affinity for the KS than for other catalytic sites.

Our results suggest a model of how ACP could first bind to the surface of the reaction chamber, close to the KS active site, and then use its flexible prosthetic arm to thread the acyl chain into the deep catalytic cleft. In a switchblade-like mechanism, the PPT arm carrying the acyl chain initially buried within the hydrophobic

Fig. 2. Structure of yeast ACP bound to the KS catalytic cleft. (A) The prosthetic PPT group (spheres) covalently attached to the ACP core (cyan) adopts an extended conformation in the fungal FAS complex. The ACP core forms a compact domain with an additional four-helix bundle (gray), rendering fungal ACP considerably larger than the bacterial counterpart (blue, shown in the same orientation). During interdomain substrate shuttling, the PTT arm might fold back on ACP (arrow), thereby inserting the acyl chain into a cavity formed by the ACP core, as observed in the isolated *E. coli* ACP structure (*16*). (B) Detailed view of PPT bound in the catalytic cleft of KS. The unbiased threefold averaged $F_{obs} - F_{calc}$ simulated annealing omit map (green) shows ACP and the phosphate and pantoic acid (PA) moieties of the PPT prosthetic group. Modeling of the additional PPT part shows that the catalytic residues of KS can easily be reached. KS1 and KS2 form the dimer to which ACP is bound.

Fig. 1. Three ACPs are stalled in each of the two reaction chambers in the heterododecameric yeast FAS complex. (A) The superposition of the yeast (green) with the *T. lanuginosus* (light blue) FAS structure reveals a very high homology (the root mean square deviation between 19725 C$_\alpha$ atoms is 1.7 Å). Although ACP is disordered in the *T. lanuginosus* crystals (*22*), it is visible in yeast (red). One type of four-helix bundle at the periphery of the central wheel could only be built in *T. lanuginosus* FAS (indicated by an asterisk). (B) Anchoring of ACP within the reaction chambers. The ACP domains, shown as red surfaces, are located between peripheral and central anchors, which are displayed as green ribbons. The flexible linkers L1 and L2, which double-tether ACP, are not seen in the electron density map but are schematically shown as dashed yellow lines. The refined portion of the PPT prosthetic group of ACP is shown in black.

Fig. 3. Mode of ACP binding to the central wheel and the catalytic cleft entrance of KS (gray). (**A**) The extended prosthetic PPT group (magenta) of ACP (cyan) reaches into the catalytic pocket of KS (asterisk). Distal to the PPT arm, ACP is flexibly tethered to the central hub (green dot) and the peripheral anchor (dashed circle) at the interior of the dome (brown). (**B**) Top and bottom views of ACP bound to the entrance cleft of KS via its core part and contacting the central wheel with the additional part. (**C**) ACP binding is mediated by complementarily charged patches, as visualized by the electrostatic surface potential of the contact areas (blue, positive; red, negative).

ACP core would flip into the KS cleft (Fig. 3A). The binding of the ACP domain to KS might even have a more active role in the threading mechanism by stabilizing the flipped-out conformation of the loop to which the PTT arm is attached. It is noteworthy that the expansion segments of the fungal FAS structure do not merely serve as passive scaffolds for the arrangement of the catalytic domains within the complex, but also guide ACP-bound substrates to the KS. A comparable mode of action might also exist for the other catalytic domains in the FAS complex.

The assignment of the ACP domain also allows us to precisely define the amino acid sequence of the flanking linkers L1 and L2, which are flexible and not visible in the electron density (Fig. 1B). An analysis of the upstream 44–amino acid–long, alanine- and proline-rich, L1 linker sequence of fungal FAS type I ACPs reveals a striking resemblance to other carrier domain tethers, such as the biotin carboxyl carrier protein (BCCP) of the acetyl–coenzyme A carboxylase (ACC) and the lipoyl domains of the pyruvate dehydrogenase (PDH) complex in *E. coli* (fig. S5) (*30*). The fact that the linkers in the FAS complex and PDH/ACC complexes have similar composition in spite of completely different folds of the ACP and the lipoyl/BCCP domains suggests that they may have evolved convergently in fungal FAS and in the other multienzymes.

The structure of the yeast FAS complex presented here shows the direct interaction of ACP with one of the key catalytic centers in a multifunctional enzyme and suggests a model for

substrate delivery that might apply to the other catalytic domains as well. These results also provide an excellent basis for future biochemical and genetic experiments in yeast, which is a well-characterized model organism.

References and Notes

1. S. W. White, J. Zheng, Y. M. Zhang, C. O. Rock, *Annu. Rev. Biochem.* **74**, 791 (2005).
2. K. J. Weissman, *Philos. Trans. A Math. Phys. Eng. Sci.* **362**, 2671 (2004).
3. R. Finking, M. A. Marahiel, *Annu. Rev. Microbiol.* **58**, 453 (2004).
4. T. Maier, S. Jenni, N. Ban, *Science* **311**, 1258 (2006).
5. S. Smith, A. Witkowski, A. K. Joshi, *Prog. Lipid Res.* **42**, 289 (2003).
6. E. Schweizer, J. Hofmann, *Microbiol. Mol. Biol. Rev.* **68**, 501 (2004).
7. F. Lynen, *Eur. J. Biochem.* **112**, 431 (1980).
8. S. Jenni, M. Leibundgut, T. Maier, N. Ban, *Science* **311**, 1263 (2006).
9. T. A. Holak, M. Nilges, J. H. Prestegard, A. M. Gronenborn, G. M. Clore, *Eur. J. Biochem.* **175**, 9 (1988).
10. G. Y. Xu et al., *Structure* **9**, 277 (2001).
11. H. C. Wong, G. Liu, Y. M. Zhang, C. O. Rock, J. Zheng, *J. Biol. Chem.* **277**, 15874 (2002).
12. G. A. Zornetzer, B. G. Fox, J. L. Markley, *Biochemistry* **45**, 5217 (2006).
13. M. P. Crump et al., *Biochemistry* **36**, 6000 (1997).
14. S. C. Findlow, C. Winsor, T. J. Simpson, J. Crosby, M. P. Crump, *Biochemistry* **42**, 8423 (2003).
15. A. Koglin et al., *Science* **312**, 273 (2006).
16. A. Roujeinikova et al., *J. Mol. Biol.* **365**, 135 (2007).
17. P. Bhaumik, M. K. Koski, T. Glumoff, J. K. Hiltunen, R. K. Wierenga, *Curr. Opin. Struct. Biol.* **15**, 621 (2005).
18. Y. M. Zhang et al., *J. Biol. Chem.* **276**, 8231 (2001).
19. Y. M. Zhang, B. Wu, J. Zheng, C. O. Rock, *J. Biol. Chem.* **278**, 52935 (2003).
20. K. D. Parris et al., *Structure* **8**, 883 (2000).
21. S. Rafi et al., *J. Biol. Chem.* **281**, 39,285 (2006).
22. S. Jenni et al., *Science* **316**, 254 (2007).
23. Materials and methods are described in the supporting material on *Science* Online.
24. H. Schuster, B. Rautenstrauss, M. Mittag, D. Stratmann, E. Schweizer, *Eur. J. Biochem.* **228**, 417 (1995).
25. F. Fichtlscherer, C. Wellein, M. Mittag, E. Schweizer, *Eur. J. Biochem.* **267**, 2666 (2000).
26. A. H. Mohamed, S. S. Chirala, N. H. Mody, W. Y. Huang, S. J. Wakil, *J. Biol. Chem.* **263**, 12315 (1988).
27. T. Schreckenbach, H. Wobser, F. Lynen, *Eur. J. Biochem.* **80**, 13 (1977).
28. G. B. Kresze, L. Steber, D. Oesterhelt, F. Lynen, *Eur. J. Biochem.* **79**, 181 (1977).
29. P. von Wettstein-Knowles, J. G. Olsen, K. A. McGuire, A. Henriksen, *FEBS J.* **273**, 695 (2006).
30. R. N. Perham, *Annu. Rev. Biochem.* **69**, 961 (2000).
31. All data were collected at the Swiss Light Source (SLS, Paul Scherrer Institute, Villigen). We are grateful to C. Schulze-Briese, S. Gutmann, E. Pohl, S. Russo, and T. Tomizaki for their outstanding support at the SLS. We thank T. Maier and J. Erzberger for critically reading the manuscript and all members of the Ban laboratory for suggestions and discussions; R. Grosse-Kunstleve, P. Afonine, and P. Adams for providing a prerelease version of the PHENIX refinement program and advice regarding structure refinement; A. Jones for a prerelease version of the program O; D. Sargent for technical assistance. Yeast cells were kindly provided by F. Hepfer of Hefe Schweiz AG. This work was supported by the Swiss National Science Foundation (SNSF) and the National Center of Excellence in Research (NCCR) Structural Biology program of the SNSF. Coordinates and structure factors of the *S. cerevisiae* structure have been deposited in the Protein Data Bank with the accession code 2UV8.

Structural Insight into Pre-B Cell Receptor Function

Alexander J. Bankovich,[1] Stefan Raunser,[6] Z. Sean Juo,[2,3,4] Thomas Walz,[6] Mark M. Davis,[1,2,5] K. Christopher Garcia[1,2,3,4]*

The pre-B cell receptor (pre-BCR) serves as a checkpoint in B cell development. In the 2.7 angstrom structure of a human pre-BCR Fab-like fragment, consisting of an antibody heavy chain (HC) paired with the surrogate light chain, the "unique regions" of VpreB and λ5 replace the complementarity-determining region 3 (CDR3) loop of an antibody light chain and appear to "probe" the HC CDR3, potentially influencing the selection of the antibody repertoire. Biochemical analysis indicates that the pre-BCR is impaired in its ability to recognize antigen, which, together with electron microscopic visualization of a pre-BCR dimer, suggests ligand-independent oligomerization as the likely signaling mechanism.

B cells traverse regulatory checkpoints to ensure that their surface-expressed antigen-specific B cell receptor (BCR) is functional yet not autoreactive (1). Upon antigen recognition by the BCR, a B cell becomes activated, expands clonally, and advances to the plasma cell stage, where it secretes soluble antibody. Before antibody genes can be transcribed, each B cell must productively rearrange one HC and one light chain (LC) gene (2). After HC expression but before the LC is rearranged to form the mature BCR, the pre-BCR is expressed (3, 4). In a functional pre-BCR, two invariant proteins, λ5 and VpreB, form a heterodimer called the surrogate light chain (SLC) that pairs with a rearranged HC (5–10).

Not all rearranged HCs can pair with the SLC to form a pre-BCR (11), but the structural basis of selection and pairing is unknown. In mice, only half of the in-frame rearranged HCs pair correctly with SLCs (11). Only those B cells that express an HC capable of pairing with the SLC undergo clonal expansion (12–15), which enriches for HCs that are capable of BCR formation (16). Humans that lack λ5 have a major block in B cell development, decreasing peripheral B cell numbers by more than 99% (17). λ5-deficient mice have a less severe phenotype, but B cell development remains substantially delayed (18).

The overall architecture of the pre-BCR is shown in Fig. 1A. The SLC/HC pair presumably contains a similar architecture of immunoglobulin (Ig)–like domains as a Fab, but the nonpolymorphic VpreB and λ5 are predicted to have divergent β-strand topologies in comparison

to canonical Ig-domain sequences (6). The N terminus of λ5 and the C terminus of VpreB have "unique regions" that share no sequence similarity to known proteins nor have a counterpart in other Ig receptors. Yet, functional studies implicate these unique regions as being essential to pre-BCR function (19–21). Both the pre-BCR and BCR signal through interaction with a heterodimeric signaling complex, Igα/Igβ, which has intracellular segments containing immunoreceptor tyrosine-based activation motifs (ITAMs) that interact with downstream adaptor and effector molecules (22). In the mature BCR, antigen engagement leads to receptor clustering and activation. However, it remains unclear whether pre-BCR signaling during normal B cell development is ligand dependent. On the one hand, several putative extracellular ligands have been identified, such as galectin-1 in humans (20) and heparin sulfate in mice (19). On the other hand, unique region–dependent, homotypic pre-BCR aggregation alone is capable of initiating downstream signaling events (21).

To gain structural insights into pre-BCR biology, we expressed and crystallized a soluble form of the human pre-BCR analogous to the Fab of an antibody (dashed box in Fig. 1A). This "Fab-like" pre-BCR was secreted from baculovirus-infected insect cells by triple co-infection of recombinant viruses encoding an ovalbumin (OVA)–specific human HC [variable region of Ig HC (V_H) and domain 1 of the constant portion of the Ig HC (C_H1)], λ5, and VpreB. We crystallized a monomeric form of the pre-BCR with the λ5 N terminus truncated by 36 residues and determined the structure to 2.7 Å (the complex with intact λ5 forms a dimer that will be discussed later). The pre-BCR adopts an overall quaternary structure that is reminiscent of a Fab. The SLC heterodimer proteins VpreB and λ5 replace the Ig LC variable and constant regions (V_L and C_L) of a Fab, respectively (Fig. 1B). The VpreB domain forms a canonical V-type Ig fold that includes β strands a to f (23). However, the VpreB polypeptide g strand is replaced by a portion of the N-terminal unique region of λ5 that completes the bottom β sheet of the V-type

Ig fold (Fig. 1C). In this arrangement, the λ5 unique region substitutes for the J region of a normal V_L that follows LC CDR3 (CDR3-L) and is involved in the V_H/V_L interface. This strand swap, which was previously predicted from homology modeling (5, 24), has no counterpart in Ig domains characterized to date. In accord with the obligate nature of SLC heterodimer formation (7), the interface between VpreB and the λ5 β strand is extensive, burying ~1000 Å^2 of surface area and forming 16 hydrogen bonds (Fig. 1D). The λ5 β strand exits from the base of VpreB and enters the λ5 constant domain. The λ5 structure is similar to a Fab C_L domain (82% identity and 1.4 Å root mean square deviation).

The region of the pre-BCR that is analogous to the antigen-binding site of antibodies contains surface-exposed loops similar to the CDR1-L and CDR2-L of a V_L (fig. S1). However, as their sequence divergence from antibodies would predict, neither loop adopts any of the known canonical LC CDR structures (25). In the pre-BCR structure, the CDR3-L of a V_L is replaced by two unconnected, frayed polypeptide chains comprising the unique regions of λ5 and VpreB (Figs. 1C and 2B and fig. S2). The proteolytic sensitivity of a portion of the λ5 unique region, which is N-terminal to the swapped β strand, suggests that this segment of the molecule is highly accessible and flexible (Fig. 1B). However, we were able to identify 11 ordered amino acids of the C-terminal unique region of VpreB in the electron density maps. This chain protrudes from the antigen-binding site, extends over the top of HC CDR3 (CDR3-H), and obscures the antigen-combining region of the HC (Fig. 2, A and B). The extended VpreB unique region structure is stabilized by a salt bridge between VpreB residue Glu[106] and HC residue Arg[59] (asterisk in Fig. 2B), although, as a result of the surface exposure, it may assume alternative conformations when complexed with different HCs [the B factors for this segment are not higher than those of the rest of the structure (table S2)].

We compared the mode of interaction between HC and SLC in the pre-BCR to that between HC and LC of all (seven) antibodies in the Protein Data Bank (PDB) database (Fig. 2C, fig. S2, and table S1) that contain the same number of residues (12) in their CDR3-H as does the HC in our pre-BCR structure. The pre-BCR exhibits a decrease of ~30% in the number of variable domain interface contacts (hydrogen bonds and van der Waals interactions) in comparison to this panel of Fabs (e.g., 53 versus 80 contacts for 1Q0X) and yet buries more surface area than most V_H/V_L pairs (915 Å^2 versus an average of 825 Å^2) (table S1). This result is due to a net loss of interactions between the SLC and HC framework residues, but there is a gain in the extent of the overall SLC/HC interface due to the SLC unique regions blanketing the HC CDR3-H (Fig. 1B). Consistent with the SLC unique regions probing CDR3-H sequences, we found more contacts (11) between the SLC unique region res-

[1]Program in Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA. [2]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA. [3]Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA 94305, USA. [4]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA. [5]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA. [6]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: kcgarcia@stanford.edu

idues and CDR3-H than between CDR3-H and CDR3-L in representative Fab structures (an average of four) (table S1).

The CDR3-H loop is extended in the pre-BCR, allowing it to simultaneously interact with SLC CDR2, the unique regions of VpreB and λ5, and other HC framework residues (Fig. 2, B to D). In contrast, in a Fab, the CDR3-H generally makes contact with only a small portion of the V_L CDR1 and CDR2 loops (fig. S3, A and B). The expansive CDR3-H interactions in the pre-BCR are enabled, by fewer structural constraints imposed on the extended and apparently flexible unique regions of VpreB and λ5 (Fig. 2, B and C, and fig. S2B), as compared to the more constrained closed-loop structure of a Fab CDR3-L. Such structural plasticity could permit unique regions to adapt to diverse CDR3-H sequences and loop lengths.

The pre-BCR appears to influence the V_H repertoire (i.e., CDR3-H sequences) of HCs in
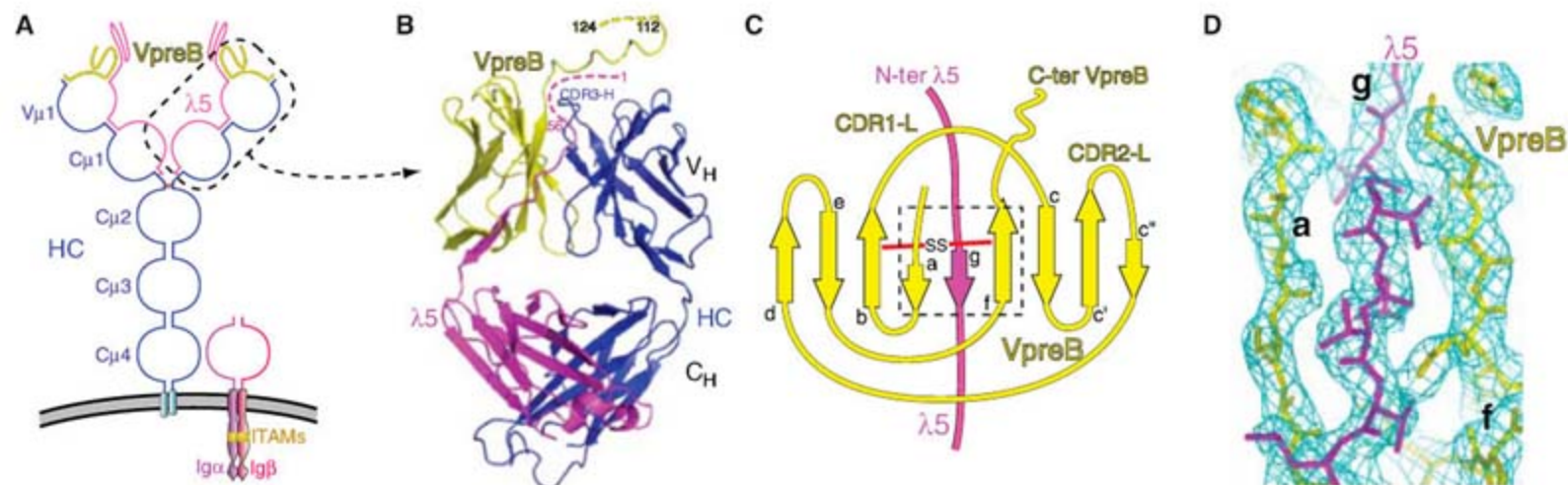
Fig. 1. Overall structure of the pre-BCR. (**A**) Cartoon representation of the pre-BCR complex with the Fab-like arm of the pre-BCR that was crystallized (dashed box). (**B**) Ribbon representation of the pre-BCR structure [dashed box in (A)]. Three protein chains are included in the model: VpreB (yellow), λ5 (magenta), and HC (blue). Missing portions of the molecule are indicated with residue numbers and dashed lines at the N terminus of λ5 and the C terminus of VpreB. (**C**) Schematic representation of the V-type Ig fold formed by VpreB and λ5. β strands are designated by arrows labeled a to g. VpreB loops that are homologous to Fab CDRs are labeled. The dashed rectangle indicates the portion of the structure shown in (D). The red line with "SS" indicates the intracellular domain disulfide bond. (**D**) 2.7 Å electron density map (σ_A-weighted $2F_o - F_c$, where F_o and F_c are the observed and calculated structure factors; contoured to 1.2σ) showing the a and f strands of VpreB and the g strand of λ5.
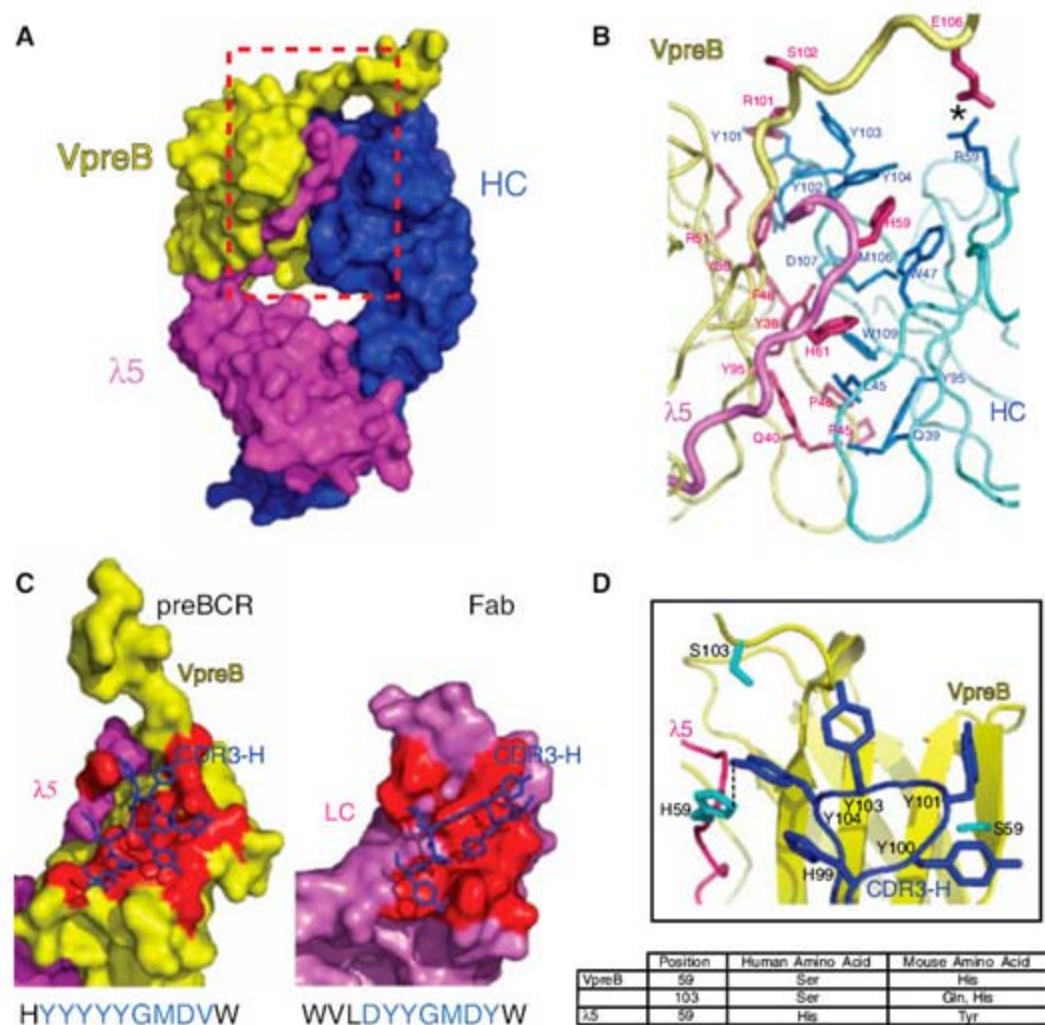
Fig. 2. The pre-BCR CDR3-H sensing site. (**A**) Surface representation of the pre-BCR structure in the same orientation as shown in Fig. 1B. The VpreB unique region exits out the top of the domain, lying over the HC antigen-combining site. (**B**) Contact residues (*29*) within the SLC/HC interface zoomed into the region demarcated by the red dashed box in (A). The main chain is depicted as ribbons with contact residue side chains as sticks. SLC, red; HC, blue. Asterisk designates the interchain salt bridge between VpreB and HC. (**C**) At left, the CDR3-H (blue sticks) from the HC on the surface of the VpreB/λ5 segment of the SLC, showing the extent of interactions between CDR3-H and the VpreB/λ5. The contact footprint between CDR3-H and SLC is highlighted in red. To the right, a similar projection of the CDR3-H on the surface of the LC of a Fab. The contact footprint between CDR3-H and LC is highlighted in red. The CDR3-H sequence is added below both structures with the region shown in sticks highlighted (blue letters). (**D**) Residues demarcating the human pre-BCR CDR3-H sensing site. CDR3-H with side chains (blue), λ5 (magenta), and VpreB (yellow) are shown. Pre-BCR side chains that vary between human and mice and contact CDR3-H are shown in cyan. The dashed line indicates a hydrogen bond. Below the structure, a table of amino acid differences between mice and humans in the region of the CDR3-H sensing site is appended.

	Position	Human Amino Acid	Mouse Amino Acid
VpreB	59	Ser	His
	103	Ser	Gln, His
λ5	59	His	Tyr

the mature B cell population (*11*). We crystallized a human pre-BCR, but the majority of genetic data on HC repertoire selection is derived from murine studies. Comparison of the HC repertoires selected by mouse and human cells reveals that murine CDR3-H sequences contain a significantly higher frequency of tyrosines and are, on average, two amino acids shorter than the human CDR3-H (*26*). The mouse and human VpreB proteins are 73% identical, and the λ5

β-strand insertion is 75% identical. Also, mice have two VpreB proteins (VpreB1 and VpreB2) that are 97% identical, differing at positions 9, 75, and 103 for VpreB1 (Leu[9], Thr[75], Gln[103]) and VpreB2 (Ser[9], Asn[75], His[103]) from human VpreB (Leu[9], Thr[75], Ser[103]). In our pre-BCR structure, positions 9 and 75 are distal to the interface with HC and are therefore unlikely to mediate pre-BCR assembly. The remaining residue, Ser[103], is located directly above the CDR3-H residue (Tyr[104]) (Fig. 2D) and is therefore close enough to play a structural role in engaging and influencing CDR3-H sequence composition.

Two pre-BCR residues make extensive contact with CDR3-H in our structure: Ser[59] in VpreB and His[59] in λ5 (Fig. 2D). VpreB Ser[59] interacts with the first two residues of CDR3-H (Tyr[100] and Tyr[101]), whereas λ5 His[59] is positioned at the C terminus of CDR3-H, and so together they bracket the CDR3-H. Along with the Ser[103] position, these residues may demarcate the boundary of a "CDR3-H sensing site" for HC repertoire selection (Fig. 2, C and D). Thus, in tandem with the many complex biological mechanisms that regulate B cell development, we suggest that the SLC unique region residues in proximity to CDR3-H may play a role in influencing the observed sequence differences between the CDR3-H repertoires in mice and humans. The CDR3-H of germline HC is the focal point for sequence diversity, with the rest of the HC and SLC being nonpolymorphic, including interfacial residues between the SLC and the HC. Thus, the site of HC discrimination is likely localized to the CDR3-H.

To address whether the pre-BCR is capable of recognizing antigen, we engineered a pre-BCR by pairing the SLC with an HC of defined antigen specificity. We used OVA-specific germline HCs from genetically limited mice (*IgH*[−/−] and *Igκ*[−/−]) containing one transgenic human HC locus and the mouse λ LC locus (*27*). The antibodies produced by these mice in response to primary immunization achieve antigen specificity exclusively through different CDR3-H sequences, showing that antigen recognition can be accomplished by diversity in only one of the six CDR loops. For pre-BCR binding experiments, we chose a panel of four OVA-specific HCs from these mice and paired them with an invariant λ1 LC or SLC to produce Fab and pre-BCR, respectively (Fig. 3, A and B). These nonsomatically mutated Fabs recognize OVA with an affinity ranging from 9 to 40 μM by means of surface plasmon resonance (SPR), whereas a hen egg-white lysozyme (HEL)–specific Fab exhibited no detectable binding (Fig. 3, B and C). For pre-BCR complexes of the same HCs and SLC, the highest affinity observed was 400 μM, essentially at the limit of detection for SPR (Fig. 3, A and C). To test the effect of LC diversity, we paired the OVA5 HC with an LC containing a point mutation in CDR3-L and found that the Fab retained binding, confirming the primacy of the CDR3-H in

C

Heavy Chain	CDR3-H Sequence	Fab K_d (μM)	preBCR K_d (μM)
OVA-1*	FRLGMT------FDIWGQG	40	>>400
OVA-5*	WRYRDWFL----FDYWGQG	9	400
OVA-6*	QRITMVRGGMG-FDYWGQG	11	NB
OVA-specific	HYYYYYG-----MDVWGQG	10	NB
HEL-1*	HVRSSSWYFYYGMDVWGQG	NB	NB
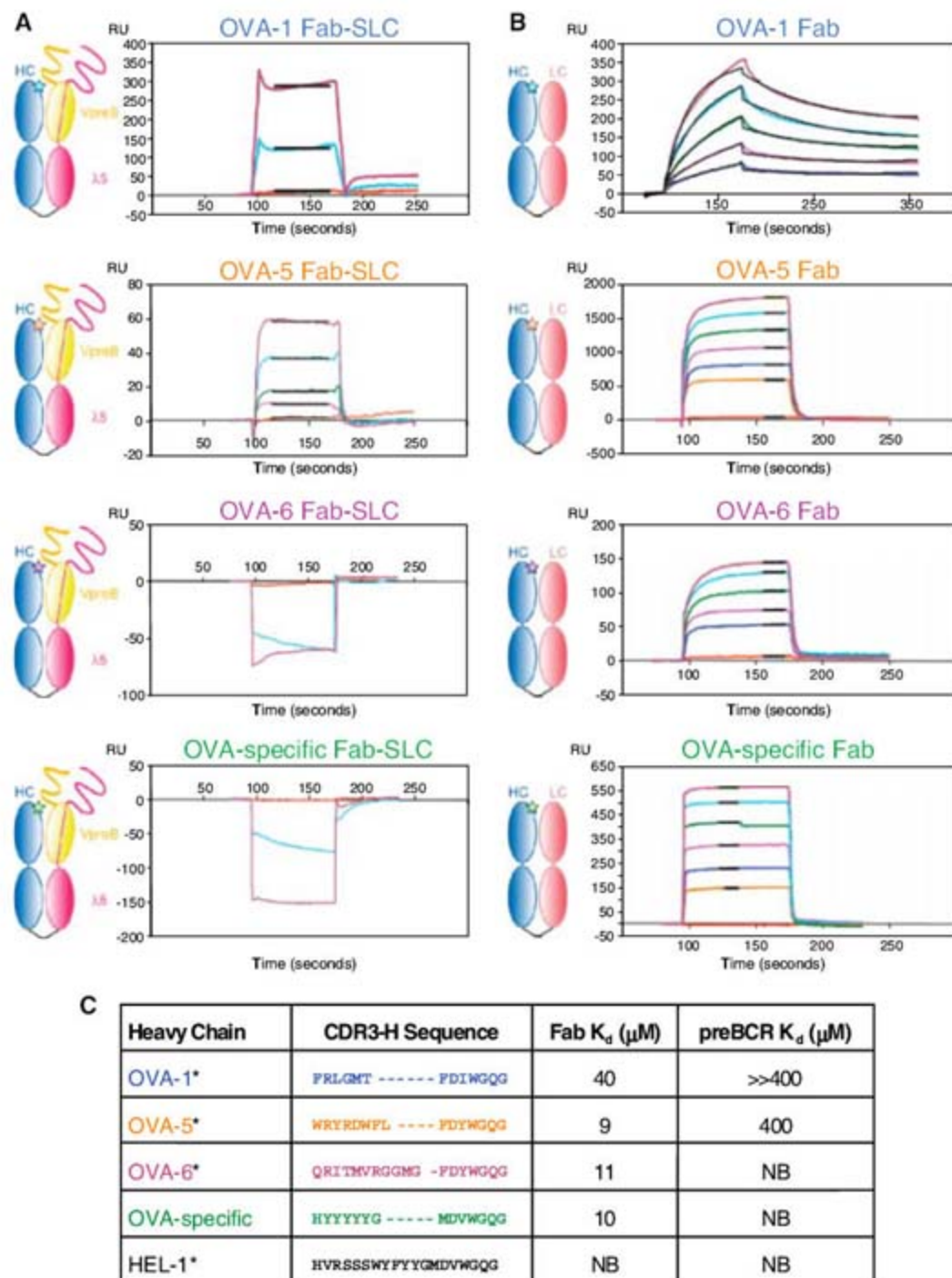
Fig. 3. Antigen-binding properties of pre-BCR versus germline Fab. (**A**) SPR sensorgrams for OVA-specific HCs in complex with SLC to form potentially "antigen-specific" pre-BCRs, which we term Fab-SLCs. Four different OVA-specific HCs are paired with SLC. The analyte concentrations shown are 400 (purple), 200 (blue), 100 (green), 50 (magenta), and 0 μM (red). Average resonance unit (RU) fits (horizontal black lines) used for steady-state affinity calculations are shown for the top two panels. CDR3-H differences are denoted by the color of the star in the schematic and the label above the sensorgram. (**B**) SPR sensorgrams for the same four OVA-specific HCs as in (A) but now in complex with the invariant λ1 LC to form a germline Fab that binds to OVA. In the top panel, 1:1 Langmuir binding affinity curve fits are shown, whereas steady-state kinetic fits (horizontal black lines) are shown for all other panels. The analyte concentrations shown are 100 (purple), 50 (light blue), 25 (green), 12.5 (magenta), 6.25 (blue), and 0 μM (red). (**C**) Summary of SPR results. Of the four Fab-SLCs tested, two did not bind, and two others exhibited binding whose affinity was >40 times less than that for the homologous Fab. K_d, dissociation constant; NB, no detectable binding. Asterisk indicates HC sequence notation [from (*27*)].

Fig. 4. The minimal signaling unit of the pre-BCR is a ligand-independent dimer. (**A**) Pre-BCR with intact SLC unique regions is a dimer. Results of size exclusion chromatography (Superdex-200, Amersham) [with retention volume (ml) on the x axis and size standards included] is shown. (**B**) SDS–polyacrylamide gel electrophoresis analysis of the gel filtration demonstrated that the dimeric peak is enriched for full-length λ5, whereas the monomeric peak is deficient in full-length λ5. Truncated λ5 comigrates with VpreB on this gel. (**C**) EM negative stain analysis of dimeric pre-BCR based on ~5400 particles. The top panels show class averages (see Methods and figs. S7 and S8), and the bottom panels show schematic interpretations based on the crys-

tal structure. Scale bar, 10 nm. (**D**) Model of the full-length cell-surface pre-BCR as a signaling active dimer that is consistent with EM and functional studies. The pre-BCR Fab-like dimers resolved by EM are boxed, and the elbow and dimer interface angles are modeled to represent the range of angles observed in EM classes.
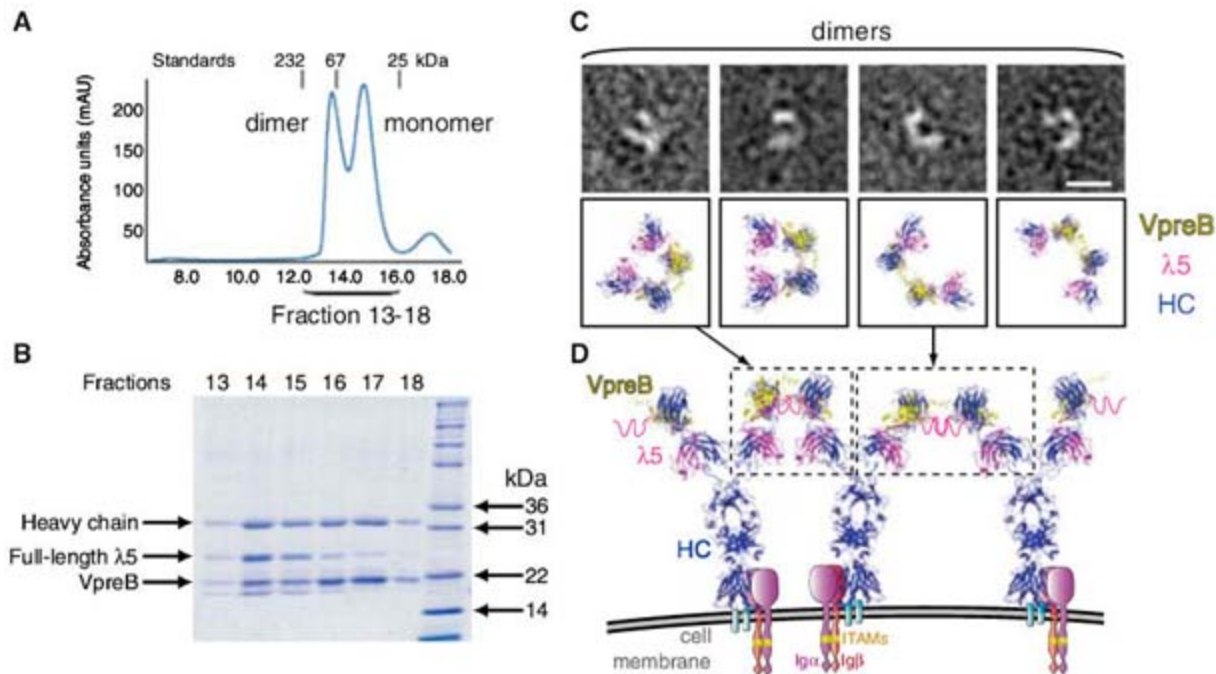
antigen recognition (fig. S4) (28). These results suggest that the pre-BCR cannot recognize protein antigens in a physiologically meaningful way, but we cannot rule out this possibility for highly abundant antigens or for small molecules. The decrease in antigen binding of the pre-BCR is consistent with the structure of the VpreB C-terminal unique region, which extends over the antigen-combining site and could partially block antigen binding (fig. S5).

Because antigen engagement does not appear to play a role in pre-BCR signaling, we revisited the concept of pre-BCR oligomerization with our recombinant pre-BCR Fab-like fragment. Size exclusion chromatography of the soluble pre-BCR yielded two peaks (Fig. 4A) with HC, λ5, and VpreB in a 1:1:1 stoichiometric ratio, which elute at the positions for a dimer and a monomer, respectively. In the monomer peak, the λ5 unique region was proteolytically truncated (Fig. 4A). Subsequent recombinant truncation of λ5 and VpreB unique regions resulted solely in monomer formation (fig. S6, A and B). Thus, perturbation of the unique regions disrupts dimerization. These data likely explain why a previously reported recombinant form of a pre-BCR Fab-like fragment, in which the VpreB and λ5 were synthetically fused to create a "single-chain" SLC, was a monomer by gel filtration. The ability of the purified recombinant pre-BCR to dimerize indicates that accessory protein(s) are not required for dimerization, and, by extension, pre-BCR signaling through multimerization can occur in a ligand-independent fashion.

Although we were unable to crystallize the dimeric form of the pre-BCR, we undertook electron microscopic (EM) studies of the pre-BCR dimer in negative stain to visualize the oligomeric architecture (Fig. 4C and figs. S7 and S8). A conservative and unbiased analysis of

both raw grids and several class averages reveals the pre-BCR dimers (Fig. 4C and figs. S7 and S8), consisting of two pre-BCR monomers (Fab-like modules) connected by a flexible hinge between the tips of the variable domains. The pre-BCR "Fab arms" in each dimer appear constrained to lying on their side, so that we observe a side-on perspective of the interdimer angle (Fig. 4C). The angle between the two pre-BCR molecules ranged from 5° to 60°, suggesting a high level of flexibility at the interface between the two domains. For each pre-BCR dimer class average, we generated a schematic model using the pre-BCR monomer crystal structure (bottom panels in Fig. 4C). We then used the λ5 truncation data, EM results, and the crystal structure to model the pre-BCR signaling complex on the cell surface (Fig. 4D). In this model, each Fab-like arm of the intact pre-BCR on the cell surface dimerizes with that of a different pre-BCR, resulting in an extended chain of self-associated pre-BCRs on the cell surface that could signal through the close proximity of their intracellular domains (Fig. 4D).

References and Notes

1. C. C. Goodnow, J. Sprent, B. Fazekas de St Groth, C. G. Vinuesa, *Nature* **435**, 590 (2005).
2. S. Tonegawa, *Nature* **302**, 575 (1983).
3. N. Sakaguchi, F. Melchers, *Nature* **324**, 579 (1986).
4. A. Kudo, F. Melchers, *EMBO J.* **6**, 2267 (1987).
5. H. Lanig, H. Bradl, H. M. Jack, *Mol. Immunol.* **40**, 1263 (2004).
6. F. Melchers et al., *Immunol. Today* **14**, 60 (1993).
7. Y. Minegishi, L. M. Hendershot, M. E. Conley, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3041 (1999).
8. T. Tsubata, M. Reth, *J. Exp. Med.* **172**, 973 (1990).
9. L. M. Hendershot, *J. Cell Biol.* **111**, 829 (1990).
10. S. Pillai, D. Baltimore, *Nature* **329**, 172 (1987).
11. E. ten Boekel, F. Melchers, A. G. Rolink, *Immunity* **7**, 357 (1997).
12. D. J. Decker, N. E. Boyle, J. A. Koziol, N. R. Klinman, *J. Immunol.* **146**, 350 (1991).
13. J. Hess et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1745 (2001).
14. H. Karasuyama et al., *Cell* **77**, 133 (1994).
15. U. Keyna, G. B. Beck-Engeser, J. Jongstra, S. E. Applequist, H. M. Jack, *J. Immunol.* **155**, 5536 (1995).
16. M. Militi, C. Schiff, M. Fougereau, C. Tonnelle *Eur. J. Immunol.* **26**, 63 (1996).
17. Y. Minegishi et al., *J. Exp. Med.* **187**, 71 (1998).
18. D. Kitamura et al., *Cell* **69**, 823 (1992).
19. H. Bradl, J. Wittmann, D. Milius, C. Vettermann, H. M. Jack, *J. Immunol.* **171**, 2338 (2003).
20. L. Gauthier, B. Rossi, F. Roux, E. Termine, C. Schiff *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13014 (2002).
21. K. Ohnishi, F. Melchers, *Nat. Immunol.* **4**, 849 (2003).
22. L. D. Wang, M. R. Clark, *Immunology* **110**, 411 (2003).
23. P. Bork, L. Holm, C. Sander, *J. Mol. Biol.* **242**, 309 (1994).
24. A. Kudo, S. Bauer, F. Melchers, *Prog. Immunol.* **7**, 339 (1989).
25. B. Al-Lazikani, A. M. Lesk, C. Chothia, *J. Mol. Biol.* **273**, 927 (1997).
26. M. Zemlin et al., *J. Mol. Biol.* **334**, 733 (2003).
27. J. L. Xu, M. M. Davis, *Immunity* **13**, 37 (2000).
28. M. M. Davis, *Semin. Immunol.* **16**, 239 (2004).
29. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
30. We thank M. Conley for providing human SLC cDNA; M. Krogsgaard for assistance with SPR; M. Boulanger, P. Strop, and X. Wang for assistance with x-ray crystallography; and M. Winslow and L. Colf for critical reading of the manuscript. This work was supported by grants from NIH, Keck Foundation, and Howard Hughes Medical Institute (K.C.G.) and the Stanford Immunology Program Training Grant (A.J.B.). The molecular EM facility at Harvard Medical School was established with a generous donation from the Giovanni Armenise Harvard Center for Structural Biology and is maintained with funds from NIH (to T.W.). Coordinates and structure factors have been deposited in the PDB (www.rcsb.org) with accession number 2H32.

Promotion of Lymphocyte Egress into Blood and Lymph by Distinct Sources of Sphingosine-1-Phosphate

Rajita Pappu,[1]* Susan R. Schwab,[2]* Ivo Cornelissen,[1] João P. Pereira[2]
Jean B. Regard,[1] Ying Xu,[2] Eric Camerer,[1] Yao-Wu Zheng,[1] Yong Huang,[3]
Jason G. Cyster,[2]† Shaun R. Coughlin[1]†

Lymphocytes require sphingosine-1-phosphate (S1P) receptor-1 to exit lymphoid organs, but the source(s) of extracellular S1P and whether S1P directly promotes egress are unknown. By using mice in which the two kinases that generate S1P were conditionally ablated, we find that plasma S1P is mainly hematopoietic in origin, with erythrocytes a major contributor, whereas lymph S1P is from a distinct radiation-resistant source. Lymphocyte egress from thymus and secondary lymphoid organs was markedly reduced in kinase-deficient mice. Restoration of S1P to plasma rescued egress to blood but not lymph, and the rescue required lymphocyte expression of S1P-receptor-1. Thus, separate sources provide S1P to plasma and lymph to help lymphocytes exit the low-S1P environment of lymphoid organs. Disruption of compartmentalized S1P signaling is a plausible mechanism by which S1P-receptor-1 agonists function as immunosuppressives.

Sphingosine-1-phosphate (S1P) functions as an extracellular signaling molecule by engaging five G protein–coupled receptors, designated $S1P_1$ to $S1P_5$, that play a key role in vascular development and lymphocyte trafficking among other processes (1). Although S1P is believed to be generated intracellularly by all cell types during sphingolipid degradation (2), the sources of extracellular S1P in vivo have not been well defined. Lymphocytes exit from lymph nodes into lymph and from the spleen into blood and are thought to exit from the thymus into blood. Studies with $S1P_1$-deficient mice revealed a necessary role for this receptor in lymphocyte egress from all three organ types (3–6). However, experiments with S1P-neutralizing antibodies and $S1P_1$ antagonists have not revealed inhibitory effects on lymphocyte egress (7–10). By contrast, synthetic $S1P_1$ agonists block lymphocyte egress and are in clinical development because of their immunosuppressive properties (11). It has been unclear how genetic $S1P_1$ deficiency can have the same outcome as increased $S1P_1$ agonism and a different outcome from $S1P_1$ antagonism. Does S1P act directly on the lymphocyte to promote egress (4–6, 12), or does it act indirectly, as pharmacological manipulations of $S1P_1$ function might suggest (7–11)?

To investigate the source of extracellular S1P and clarify its role in egress, we generated mice deficient in the two enzymes responsible for S1P production, sphingosine kinase-1 (Sphk1) and sphingosine kinase-2 (Sphk2) (2, 13). Because deletion of either kinase alone did not have a profound effect on circulatory S1P and combined deletion led to embryonic lethality at mid-gestation (14–17), we used a conditional gene deletion approach (fig. S1, A to E). $Sphk1^{f/-}Sphk2^{-/-}$ pups carrying the Mx1-Cre transgene were treated 3 to 5 days after birth with polyinosine polycytidylic acid (pI-pC) to activate the Mx1 promoter and Cre expression. Such induction of Mx1-Cre causes efficient excision of floxed alleles in hematopoietic cells, vascular endothelium, and liver, among other organs (fig. S1F) (18, 19). We further established that Mx1-Cre is active in lymphatic endothelium (fig. S1, G and H). pI-pC–treated $Sphk1^{f/-}Sphk2^{-/-}$ Mx1-Cre mice, hereafter designated "Sphk-deficient" mice, survived to adulthood and were indistinguishable from littermates in appearance.

By liquid chromatography followed by tandem mass spectrometry (LC/MS/MS), S1P was undetectable in plasma and lymph of Sphk-deficient mice, whereas the concentrations of S1P in plasma and lymph of controls were in the micromolar and the 100-nanomolar range, respectively (Fig. 1A). Similar differences were obtained in a bioassay of S1P activity (fig. S1I). Furthermore, flow cytometric analysis of the lymphocytes that could be identified in blood and lymph of the Sphk-deficient mice revealed high surface $S1P_1$ levels compared with controls, where receptor expression was undetectable (Fig. 1B). $S1P_1$ is highly sensitive to ligand-induced internalization, and thymocytes down-modulate $S1P_1$ almost completely when exposed to 1 nM S1P (12). The high surface $S1P_1$ on circulating lymphocytes in the Sphk-deficient mice is consistent with a lack of circulatory S1P. $S1P_1$ levels on cells from the thymus, spleen, and lymph nodes of Sphk-deficient mice overlapped with $S1P_1$ levels on blood and lymph lymphocytes from these mice (Fig. 1C and fig. S1J), suggesting that extracellular S1P is equally low in all sites.

The number of T cells in blood and lymph of Sphk-deficient mice was markedly reduced, whereas spleen and lymph node T cell numbers were similar compared to controls (Fig. 2, A to

[1]Cardiovascular Research Institute, University of California, San Francisco, 600 16th Street S472D, San Francisco, CA 94143–2240, USA. [2]Howard Hughes Medical Institute (HHMI) and Department of Microbiology and Immunology, University of California, San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143–0414, USA. [3]Drug Studies Unit, Department of Biopharmaceutical Sciences, University of California, San Francisco, 296 Lawrence Street, South San Francisco, CA 94080, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: Shaun.Coughlin@ucsf.edu (S.R.C.); Jason.Cyster@ucsf.edu (J.G.C.)

Fig. 1. Sphk-deficient mice have no detectable S1P in plasma or lymph. (**A**) LC/MS/MS quantification of plasma and lymph S1P. The limit of detection in this assay was about 50 nM (5×10^{-2} on ordinate). Δ indicates $Sphk2^{-/-}$ mice in which one Sphk1 allele was null and one Sphk1 allele was excised; F indicates $Sphk2^{-/-}$ littermate controls in which at least one functional Sphk1 allele was present. Both Δ and F mice were treated with pI-pC 3 to 5 days after birth. (**B**) $S1P_1$ levels on $CD4^+CD62L^{hi}$ cells from blood or lymph of the indicated mice, measured by flow cytometry. Shaded histograms show staining of blood $CD4^+CD62L^{hi}$ cells with a control antibody. Each plot is representative of at least three experiments. (**C**) $S1P_1$ levels on $CD4^+CD62L^{hi}$ cells from the indicated organs. Each plot is representative of at least three experiments.

D). Furthermore, there was an accumulation of mature T cells in the thymus without apparent changes in other thymic subsets (Fig. 2E). These observations indicate that Sphk is required for T cell egress from peripheral lymphoid organs and from the thymus. The presence of close-to-normal T cell numbers in peripheral lymphoid organs despite the defect in thymic egress could be due to the persistence of some Sphk-expressing cells up to a week or more after birth. B cell counts in blood and lymph of Sphk-deficient mice were also significantly lower than those in controls (Fig. 2, A and C), although B

cell numbers in the lymphoid organs of Sphk-deficient mice appeared normal (Fig. 2, B and D, and fig. S2A). B cell egress from the bone marrow, which occurs in the absence of $S1P_1$ (4), presumably provided a source of peripheral B cells that entered but failed to exit lymphoid organs in Sphk-deficient mice. It has been suggested that $S1P_1$, and thus possibly S1P, might be needed for normal lymphocyte development rather than for egress (11). However, short-term adoptive transfer experiments indicate that the defective egress reflects a proximal requirement for S1P (fig. S2B).

To determine whether the source of S1P that promotes lymphocyte egress is hematopoietic in origin, we performed bone marrow reconstitution studies. In lethally irradiated wild-type mice transplanted with Sphk-deficient marrow, plasma S1P levels were reduced by 10-fold compared with controls, although lymph S1P was normal (Fig. 3A and fig. S3A). The remaining plasma S1P concentration was above the low nanomolar range sufficient to down-modulate lymphocyte $S1P_1$ in vitro (12), and $S1P_1$ on blood lymphocytes remained low (Fig. 3B). Lymphocyte counts in both blood and lymph were in the nor-
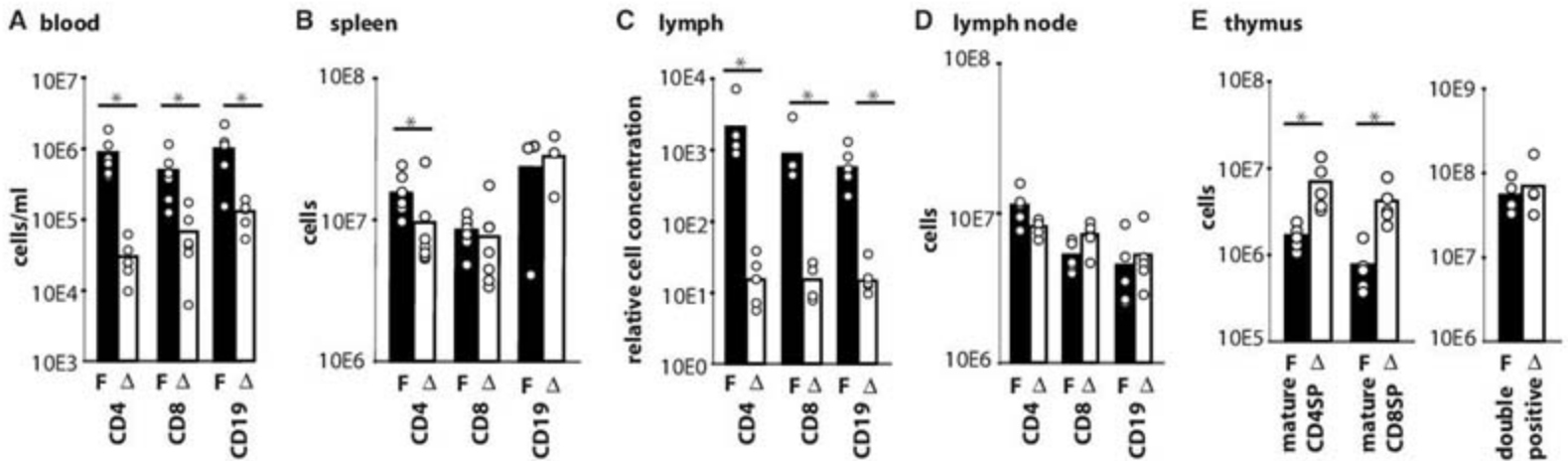
Fig. 2. Lymphocytes fail to egress from thymus and secondary lymphoid organs in Sphk-deficient mice. (**A** to **E**) Cell numbers in the indicated organs or fluids from control or Sphk-deficient mice. Lymph node numbers are for the mesenteric lymph nodes. All enumerated cells (except double-positive thymocytes) were CD62L[hi]; additionally, splenic CD19[+] cells were IgD[hi] and AA4.1[lo], and mature CD4 and CD8 single-positive (SP) thymocytes were CD69[lo]. Asterisks indicate $P \le 0.05$.
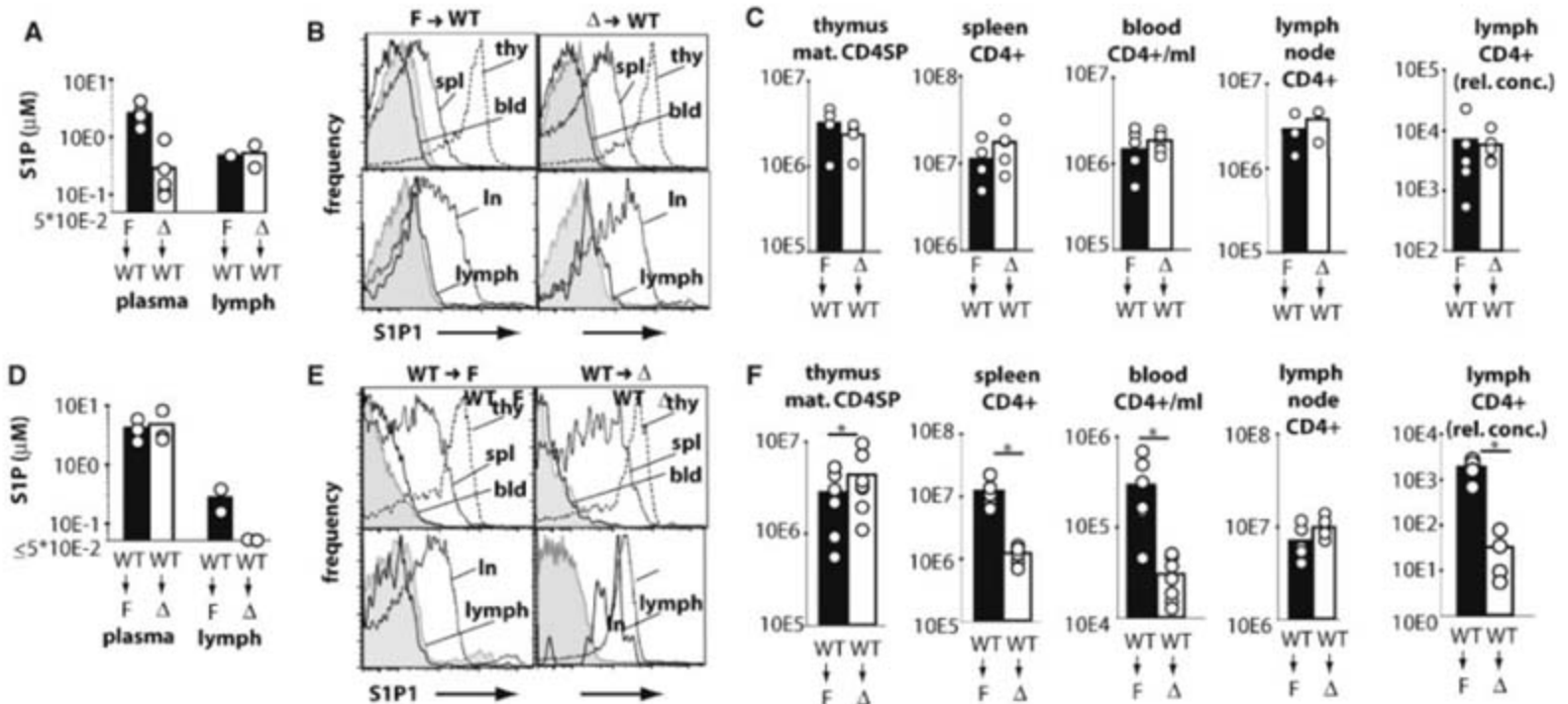
Fig. 3. A hematopoietic source supplies most plasma S1P, whereas a radiation-resistant source supplies lymph S1P. (**A** to **C**) Lethally irradiated wild-type (WT) mice were reconstituted with Sphk-deficient or littermate control bone marrow. (**D** to **F**) Lethally irradiated Sphk-deficient or control mice were reconstituted with wild-type bone marrow. [(A) and (D)] LC/MS/MS quantification of plasma and lymph S1P in the indicated mice. [(B) and (E)] $S1P_1$ levels on CD4[+]CD62L[hi] cells from the indicated organs and fluids, measured by flow cytometry. Each plot is representative of at least three experiments. Experiments shown in bottom graphs in (B) were done on the same day and used the same background control. [(C) and (F)] Cell numbers in the indicated organs and fluids. Lymph node numbers are for the mesenteric lymph nodes. All enumerated cells were CD62L[hi]; additionally mature CD4 single-positive thymocytes were CD69[lo]. For lymph, relative concentration (rel. conc.) was determined as described in (13). Asterisks indicate $P \le 0.05$.

mal range, and mature T cells did not accumulate in the thymus (Fig. 3C and fig. S3, B and C). These data suggest that S1P in plasma is supplied mainly by hematopoietic cells and that S1P in lymph is supplied by an independent source. They also suggest that radiation-resistant cells generate sufficient S1P to support lymphocyte egress into blood and lymph.

In lethally irradiated Sphk-deficient mice reconstituted with wild-type bone marrow, the S1P level in plasma was restored to the normal range, but S1P in lymph remained undetectable (Fig. 3, D and E, and fig. S3D). Egress from the thymus was largely restored: Mature CD4 single-

positive thymocytes accumulated only ~50% above control (Fig. 3F) compared with a greater than fourfold accumulation in Sphk-deficient mice (Fig. 2E). Although some studies have suggested thymic egress can occur via lymphatics (20–22), this finding supports the view that exit can occur into the blood. T and B cell numbers were dramatically reduced in the lymph (Fig. 3F and fig. S3, E and F), suggesting a persistent block in lymphocyte egress from lymph nodes. Furthermore, there was a substantial decrease in the number of T and B cells in the spleen and blood and an increase in the lymph nodes (Fig. 3F and fig. S3, E and F). This suggested that splenic

lymphocytes were exiting into the blood, traveling to lymph nodes, and then failing to exit, resulting in a general block in recirculation. Because the number of CD4[+] T cells in the entire lymph node compartment is substantially larger than the number in the spleen, their fractional decrease in spleen was easier to detect than their fractional increase in lymph nodes. These data are again consistent with a model in which plasma S1P is supplied predominantly by hematopoietic cells and lymph S1P is derived from a separate source(s). Moreover, although the hematopoietic source is sufficient to support lymphocyte egress from thymus and spleen, only the separate Mx1-
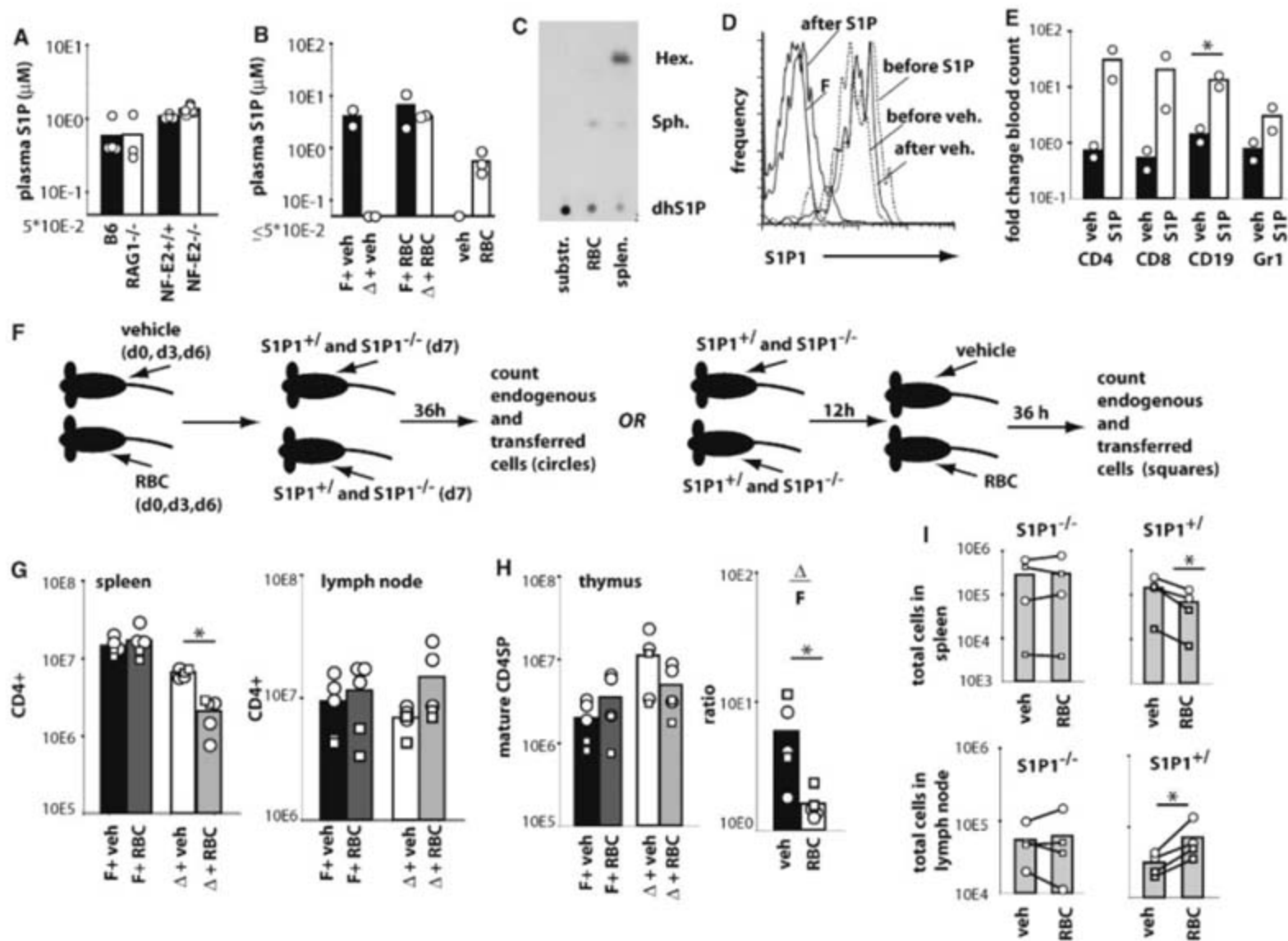
Fig. 4. Plasma S1P promotes lymphocyte egress into the blood in an $S1P_1$-dependent manner. (**A**) LC/MS/MS quantification of plasma S1P from the indicated mice. (**B**) Sphk-deficient or control mice received 0.35 ml of packed RBCs (about 2×10^9 cells) or vehicle 36 hours before analysis. Plasma S1P, as well as S1P in the RBC preparation, was quantified by LC/MS/MS. (**C**) S1P lyase activity. Tritiated dihydroS1P (dhS1P) was incubated with lysates of RBCs, splenocytes (splen.), or homogenization buffer alone (substr.). The hexadecanal (Hex) product was separated by thin-layer chromatography and visualized by autoradiography. Some dephosphorylation of dhS1P to dihydrosphingosine (Sph) also occurred. Results are representative of three experiments. (**D** and **E**) Sphk-deficient mice were infused intrajugularly with S1P or vehicle for 4 hours. (D) $S1P_1$ levels on blood CD4[+]CD62L[hi] cells, measured by flow cytometry. Plot is representative of two experiments. (E) Blood counts of the indicated cell types were measured before and after

infusion, and the fold change is plotted. All enumerated lymphocytes were CD62L[hi]. (**F**) Experimental design for (G) to (I). (**G** and **H**) Cell numbers in the indicated organs of RBC-transfused or control mice. Enumerated cells were CD4[+]CD62L[hi]; in addition, mature CD4SP thymocytes were CD69[lo]. In (H), the left plot shows the number of mature CD4SP thymocytes, and the right plot shows the ratio of the number of mature CD4SP thymocytes in the thymus of a Sphk-deficient mouse to the number in its identically treated Sphk-sufficient littermate control; each point represents an individual pair. (**I**) Mature CD4SP $S1P_1^{-/-}$ and $S1P_1^{+/}$ thymocytes were cotransferred into Sphk-deficient mice that also received RBC or vehicle. Transferred CD4[+]CD62L[hi] cells were enumerated in spleen and lymph nodes after 36 hours. Four experiments are shown; lines connect pairs of mice in the same experiment. Circles and squares in (G) to (I) represent individual mice as indicated in (F). Asterisks indicate $P \leq 0.05$.

Cre–sensitive, radiation-resistant source can support lymphocyte egress from lymph nodes.

Previous studies have suggested that platelets might be an important source of plasma S1P (23–25). However, NF-E2–deficient mice had normal plasma S1P concentrations (Fig. 4A) despite having virtually no circulating platelets (26). Analysis of RAG1-deficient mice that lack T and B cells also did not reveal any reduction in plasma S1P (Fig. 4A). These observations and our finding that plasma and lymph S1P are separately maintained led us to investigate whether red blood cells (RBCs) are a source of plasma S1P. Wild-type RBCs (fig. S4, A to F) were transferred to Sphk-deficient mice in amounts sufficient to represent at least 20% of circulating RBCs in the recipients. Within 36 hours of wild-type RBC transfer, plasma S1P in Sphk-deficient mice was restored (Fig. 4B and fig. S4G). RBCs have abundant sphingosine kinase activity (fig. S4H), as expected (27). Moreover, although all cell types except platelets have been thought to express the S1P-degrading enzyme S1P lyase (2, 24), we observed that RBCs contain minimal lyase activity (Fig. 4C). This low lyase activity, together with their abundance, suggests a role for RBCs as the main source of plasma S1P and accords with the inability of lyase inhibitors to alter blood S1P levels (12).

The ability of hematopoietic cell–derived S1P to support lymphocyte egress from the thymus and spleen and the evidence that RBCs are the major hematopoietic S1P source suggested that thymocytes and spleen cells egress in direct response to plasma S1P. Thus, we asked whether providing S1P in the circulation might be sufficient to restore egress of cells from lymphoid tissues into blood. S1P infusion into Sphk-deficient mice down-modulated $S1P_1$ surface expression on blood T cells and indeed led to an increase in the number of B and T cells in the blood within 4 hours (Fig. 4, D and E). To perform longer-term experiments, we used RBC transfusion. Three RBC transfusions over 9 days led to an increase in plasma but not lymph S1P (fig. S4I). RBC transfusion by this protocol or transfusion once with analysis at 36 hours produced a decrease in T and B cell numbers in the spleen and an accumulation in the lymph nodes of Sphk-deficient mice (Fig. 4, F and G). There was also a reduction in the number of mature T cells in the thymus that was most obvious when the data were analyzed as the ratio of counts from identically treated littermate pairs (Fig. 4H). To test whether the RBC transfusions

act by providing S1P to stimulate $S1P_1$ on the lymphocyte or by some indirect mechanism, we cotransferred fluorescently labeled $S1P_1^{-/-}$ and $^{+/}$ (+/+ or +/−) lymphocytes into Sphk-deficient mice, either 1 day after the third RBC transfusion in the 9-day protocol or 12 hours before the single RBC transfusion in the 36-hour protocol (Fig. 4F). RBC transfusion reduced the number of $S1P_1^{+/}$ cells in the spleen and increased their number in lymph nodes, but the numbers of $S1P_1^{-/-}$ cells in the spleen and lymph nodes were unaffected (Fig. 4I). These findings suggest that RBC-derived plasma S1P is able to promote lymphocyte egress from the spleen in a manner that depends on lymphocyte-intrinsic expression of $S1P_1$.

The above findings identify a previously unknown, endocrine-like function for RBCs as a major producer of plasma S1P. We also make the unexpected finding that plasma and lymph S1P are maintained as separate compartments. Albumin and high-density lipoproteins, major S1P carriers, are thought to be supplied to lymph from transudated plasma and are present in lymph at concentrations within a fewfold of those in plasma (28). Our data suggest that S1P is removed from transudated plasma and "resupplied" to lymph by a distinct source, perhaps lymphatic endothelial cells, which we have shown are a radiation-resistant Mx1-Cre target (fig. S1, and H). This may be part of a general mechanism that keeps S1P abundance in tissues low, ensuring S1P-receptor–expressing cells are sensitive to local increases in S1P, as may occur at a site of tissue injury or inflammation (2, 24).

Taken together, our data are consistent with a model in which egress from thymus, spleen, and lymph nodes is promoted by S1P produced by radiation-resistant cells at egress sites, as well as, in the case of thymus and spleen, by plasma S1P. S1P at exit sites enables egress by acting on $S1P_1$ expressed by lymphocytes. In accord with studies of $S1P_1$-deficient mice (4–6), this model predicts that antagonism of $S1P_1$ should inhibit egress; the inability of the antagonists reported thus far to do so (7–10) is likely due to incomplete receptor blockade or ligand neutralization, because our studies of wild-type mice reconstituted with Sphk-deficient marrow revealed that even when plasma S1P levels were reduced by 95%, lymphocyte egress occurred normally. This model also predicts that activation and down-regulation of lymphocyte $S1P_1$ within lymphoid tissue, either by elevating tissue S1P by inhibition of lyase activity (12) or by direct activation of $S1P_1$

with the lyase-insensitive agonist FTY720 (11), would disrupt compartmentalization of S1P signaling, which is normally low in lymphoid organ interstitium and high at exit sites, and inhibit egress. Thus, this model accommodates the paradox that $S1P_1$ deficiency and $S1P_1$ agonism can both yield defective lymphocyte egress.

References and Notes
1. T. Hla, Pharmacol. Res. 47, 401 (2003).
2. S. Spiegel, S. Milstien, Nat. Rev. Mol. Cell Biol. 4, 397 (2003).
3. M. L. Allende, R. L. Proia, Biochim. Biophys. Acta 1582, 222 (2002).
4. M. Matloubian et al., Nature 427, 355 (2004).
5. M. L. Allende, J. L. Dreier, S. Mandala, R. L. Proia, J. Biol. Chem. 279, 15396 (2004).
6. J. G. Cyster, Annu. Rev. Immunol. 23, 127 (2005).
7. S. H. Wei et al., Nat. Immunol. 6, 1228 (2005).
8. M. G. Sanna et al., Nat. Chem. Biol. 2, 434 (2006).
9. B. Visentin et al., Cancer Cell 9, 225 (2006).
10. F. W. Foss Jr. et al., Bioorg. Med. Chem. 15, 663 (2007).
11. H. Rosen, E. J. Goetzl, Nat. Rev. Immunol. 5, 560 (2005).
12. S. R. Schwab et al., Science 309, 1735 (2005).
13. Materials and methods are available as supporting material on Science Online.
14. M. L. Allende et al., J. Biol. Chem. 279, 52487 (2004).
15. B. Zemann et al., Blood 107, 1454 (2006).
16. K. Mizugishi et al., Mol. Cell. Biol. 25, 11113 (2005).
17. Y. Kharel et al., J. Biol. Chem. 280, 36865 (2005).
18. R. Kuhn, F. Schwenk, M. Aguet, K. Rajewsky, Science 269, 1427 (1995).
19. C. E. Leuker, M. Labow, W. Muller, N. Wagner J. Exp. Med. 193, 755 (2001).
20. S. Kato, Microsc. Res. Tech. 38, 287 (1997).
21. M. Kotani, M. Kawakita, M. Fukanogi, A. Yamashita, K. Seiki, Okajimas Folia Anat. Jpn. 43, 61 (1967).
22. T. Ushiki, Cell Tissue Res. 244, 285 (1986).
23. Y. Yatomi, S. Yamamura, F. Ruan, Y. Igarashi J. Biol. Chem. 272, 5291 (1997).
24. Y. Yatomi, Y. Ozaki, T. Ohmori, Y. Igarashi, Prostaglandins Other Lipid Mediat. 64, 107 (2001).
25. S. Aoki et al., J. Biochem. (Tokyo) 138, 47 (2005).
26. R. A. Shivdasani et al., Cell 81, 695 (1995).
27. L. Yang, Y. Yatomi, Y. Miura, K. Satoh, Y. Ozaki, Br. J. Haematol. 107, 282 (1999).
28. C. H. Sloop, L. Dory, P. S. Roheim, J. Lipid Res. 28, 225 (1987).
29. We thank T. Okada, L. Shiow, and S. Glantz for discussions. R.P. was supported by NIH training grant HL07731. S.R.S. is a Robert Black fellow of the Damon Runyon Cancer Research Foundation. J.G.C. is a HHMI Investigator. This work was supported in part by grants from the NIH to S.R.C. and J.G.C.

Vertical Gel Casting System

The omniPAGE is an easy-to-use, leak-proof mini vertical gel casting system for protein polyacrylamide gel electrophoresis (PAGE) separations and electroblotting applications. The unit is compatible with all major brands of precast gels and offers important features and benefits to users. A novel, quick-release clamp system provides a rapid, simple method for casting mini vertical gels. The patented sliding-clamp mechanism means that the sealing and casting of gels is achieved by sliding two clamps on either side of the glass plates, a process that takes seconds. To cast the gel, the glass plates are simply inserted into the unit and clamps slid across to seal. The insert is then transferred to the casting base and secured down. The gel solution is loaded and once it has set, the module is placed into the tank for running. This makes gel casting a simple four-step procedure that is considerably faster than in systems that make use of complicated, unreliable screw mechanisms or require the transfer of glass plates.

Cleaver Scientific For information +44 (0) 1788 565 300 www.cleaverscientific.com

Microdissection Instrument

The Arcturus laser capture microdissection instrument is built on the Nikon TE2000U inverted research grade microscope and combines infrared (IR) laser capture microdissection (LCM) and ultraviolet (UV) laser cutting in a single modular instrument, offering researchers superior speed, precision, and flexibility. Because it makes use of the Nikon TE2000U, the Arcturus is completely modular and upgradeable, allowing the system to expand as research requirements grow and microdissection technologies emerge. The instrument features exclusive LCM technology and automated image archiving that takes the guesswork out of microdissection. Optional high-intensity halogen illumination enables phase contrast and differential interference contrast, making the system suitable for live-cell applications. Additional options enable system modification for alternate applications and potential add-ons such as optical trap laser tweezers for high-precision micromanipulation of microscopic particles or an additional camera for high-resolution imaging. **Molecular Devices** For information 408-747-3533 www.moleculardevices.com

Refrigerated Solvent Condenser

The Speed Trap refrigerated solvent condenser for vacuum concentrators is designed to improve the performance of any vacuum concentration system through the use of fluoropolymer components that provide enhanced solvent resistance and a high-safety plastic-coated solvent collection vessel. The Speed Trap features a novel, efficient cooled condenser coil. The cold condenser coils are suspended directly in the vapor path. Solvent vapors condense directly onto the coils and run off into the collection vessel below without freezing, even when water is used. Users can readily see the solvents in the trap, and emptying requires no defrosting. The plastic-coated collection vessel is removed with a simple quarter turn, allowing users to empty the cold trap quickly and safely.
Genevac For information 845-267-2211 www.genevac.com

Biomarker and Drug Imaging Software

New flexImaging 2.0 software is for acquisition and evaluation of matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) and TOF/TOF imaging data. MALDI imaging of tissue samples with flexImaging 2.0 allows color-coded visualization of the distribution of biomarkers or absorbed drugs and their metabolites. Integrating multivariate statistical classifications such as variance ranking, flexImaging 2.0 now provides "Class Imaging" as an extension to the established "Mass Imaging." Class imaging allows the classification of tissue types and determination of the class membership of comparable tissues from other individuals or animals used in drug development. With this software, the statistical biomarker and drug and metabolite data obtained with the Flex series of MALDI-TOF and TOF/TOF mass spectrometers can now be merged easily with molecular images for in-depth investigation of the spatial distribution of relevant marker candidates in tissues.
Bruker Daltonics For information 978-667-9580 www.bdal.com

Microfluidic Workstation

The BioRaptr FRD microfluidic workstation is suitable for assay miniaturization applications in drug discovery, genomics, and proteomics. The BioRaptr non-contact, high-precision dispenser delivers increased capacity and speed along with reduced reagent and consumable costs. The instrument is designed so the plate moves while the pipetting head is stationary, creating a shorter fluid path, for high precision with low dead volume. Flexible programming is easily done through an Excel spreadsheet and allows a variety of reagents to be dispensed into any well at a range of volumes. The workstation can be combined with Beckman Coulter's Biomek and Paradigm systems for high-throughput applications such as compound screening.
Beckman Coulter For information 714-993-8955 www.beckmancoulter.com

Dual-Mode Plate Readers

The Lucy 2 and Lucy 3 are multi-mode microplate readers with an emphasis on luminescence and absorbance capabilities. Lucy 2 operation is controlled by a personal computer. Lucy 3 features an onboard computer, display, and keyboard for stand-alone function with comprehensive evaluation software. The photon-counting detection system features high sensitivity and dynamic range. Two fast reagent injectors facilitate both flash and glow-type luminescence assays. In addition, absorbance reading in the visible range can be performed. A high-capacity filter wheel of up to 15 filters provides flexibility when switching from luminescence to absorbance mode for enzyme-linked immunosorbent assay or similar applications. The instrument features a fast flash kinetic mode (with kinetics in the microsecond range), a luminometric range from 300 nm to 650 nm, and a photometric range from 405 nm to 650 nm.
Anthos For information +44 (0) 1223 427812 www.anthos-labtec.com

ASSISTANT PROFESSOR
Infectious Disease Faculty Position
University of Connecticut

The University of Connecticut invites applications
for a tenure-track position in the Department of
Pathobiology and Veterinary Science. This position
requires an advanced degree (D.V.M./Ph.D., Ph.D.
or equivalent) and postdoctoral experience. Individ-
uals with expertise in microbial pathogenesis, im-
munology, or pathology of infectious diseases are
encouraged to apply. The successful candidate is ex-
pected to establish/maintain an extramurally funded
research program that will complement the existing
programs within the Department and to contribute
to the departmental undergraduate and graduate
teaching and service missions. The Department of
Pathobiology and Veterinary Science encompasses a
diverse group of disciplines including pathology,
virology, immunology, and bacteriology.

Review of applications will begin immediately and
continue until a successful candidate is identified. Ap-
plicants should electronically submit curriculum vitae
and a statement of research interest and teaching phi-
losophy, as well as arrange for three letters of rec-
ommendation to be sent to: **Mrs. Maria Chartier,
Faculty Search Committee Coordinator, Univer-
sity of Connecticut, Department of Pathobiology
and Veterinary Science, 61 North Eagleville Road,
Unit 3089, Storrs, CT 06268-3089; e-mail:
maria.chartier@uconn.edu.** Information about the
Department and faculty may be found at the College
website: http://www.patho.uconn.edu. (Search #
2007059.)

*The University of Connecticut is an Equal Opportunity
Employer and Program Provider and encourages applications
from underrepresented groups including minorities, women, and
people with disabilities.*

FACULTY POSITION in CELL BIOLOGY

The Department of Cell Biology of the University
of Texas Southwestern Medical Center at Dallas
seeks applications for a tenure-track position at the
level of **ASSISTANT PROFESSOR.** Applicants
must have a Ph.D. or M.D. degree, at least two
years of postdoctoral training, and experience teach-
ing human anatomy. Candidates are expected to de-
velop an independent research program. Preference
will be given to candidates with research programs in
one of the following general research areas: cell biol-
ogy, neuroscience, or functional morphology. The
successful candidate will be joining a highly inter-
active and collegial teaching faculty. Applicants
should e-mail their curriculum vitae, the names of
three references, and a brief description of their re-
search goals to the attention of **Dr. Richard Anderson**
at **e-mail: cb.recruitment@utsouthwestern.edu.**
For more information go to **website: http://
www.utsouthwestern.edu/utsw/cda/dept25128/
files/34664.html.** *The University of Texas Southwestern
Medical Center is an Equal Opportunity/Affirmative Action
Employer.*

ASSISTANT or ASSOCIATE PROFESSOR
Pharmacology

Tenure-track, full-time position that is 100 percent
fully funded by the Institution. Applicants must have
a Ph.D. in pharmacology or a related biomedical
discipline, or a D.O. or an M.D. degree, and at least
two years of postgraduate, productive academic
experience. Candidates should send complete curric-
ulum vitae, including reference names, to: **Human
Resources, Kirksville College of Osteopathic
Medicine, A.T. Still University of Health Sci-
ences, 800 West Jefferson Street, Kirksville, MO
63501.** The target date for filling the position is
August 1, 2007.

*The A.T. Still University of Health Sciences (ATSU) is an
Affirmative Action/Equal Opportunity Employer. ATSU
does not discriminate on the basis of race, color, religion,
national origin, gender, age, disability, or status as a Vietnam-
era veteran in its hiring and employment practices.*

The Mayo Clinic Section of Scientific Publications
seeks a full-time **MANUSCRIPT EDITOR.** The Sec-
tion has offices at Rochester, Minnesota, Jacksonville,
Florida, and Scottsdale, Arizona, and is one of the
largest academic departments of biomedical editing
in the United States. This position supplements the
current seven biomedical editors and 20 editorial as-
sistants and proofreaders. Candidates should have
significant experience in editing peer-reviewed manu-
scripts, book chapters, and NIH grants. A relevant
Ph.D. degree or substantial academic editing expe-
rience is required. Candidates will be expected to
furnish examples of their work so that the edited man-
uscript can be compared with the unedited copy.
Successful applicants will hold a Faculty position at
Mayo Clinic College of Medicine (level commensu-
rate with experience). Excellent salary and benefits.

Send curriculum vitae to:

**Rosemary Perry
Administrative Assistant
Scientific Publications
Plummer Building, S-10
200 First Street S.W.
Rochester, MN 55905**
E-mail: perry.rosemary@mayo.edu

*Mayo Foundation is an Affirmative Action and Equal Op-
portunity Employer and Educator.*

The Department of Surgery at the University of
Pennsylvania's School of Medicine seeks candidates
for an **ASSISTANT PROFESSOR** position in the
nontenure research track. The successful applicant
will have experience in the field of islet transplanta-
tion and diabetes research. Responsibilities include
assisting in the development of a nationally funded
program to study the role of novel approaches for
the treatment of diabetes, as well as contributing to
ongoing research projects in the fields of human islet
isolation and transplantation and islet physiology
and biochemistry. This is a research-only position,
with no teaching or clinical responsibilities. Appli-
cants must have an M.D. and/or Ph.D degree.

Candidates must have three years or more of
postdoctoral and faculty-level or equivalent experi-
ence. Preference will be given to those with a strong
history of successful research in areas including, but
not limited to, transplant immunobiology, pancreatic
islet isolation, and physiology of glucose-stimulated
insulin secretion and beta cell regeneration.

Please submit curriculum vitae, a letter of interest,
and references to:

**Ali Naji, M.D., Ph.D.
J. William White Professor of Surgical Research
c/o Peter Atherton
University of Pennsylvania School of Medicine
3400 Spruce Street, 4029 Maloney
Philadelphia, PA 19104-4283**
E-mail: athertop@uphs.upenn.edu

Please refer to ad # RschAsstProf_Transplant_0707.

*The University of Pennsylvania is an Equal Opportunity,
Affirmative Action Employer. Women and minority candidates
are strongly encouraged to apply.*

ASSOCIATE RESEARCH SCIENTIST and POSTDOCTORAL RESEARCH POSITIONS,

Department of Medicine - Pulmonary. Positions
available for highly motivated individuals with Ph.D.
or M.D. NIH-funded studies involve laboratory re-
search in mouse models for lung cancer and the role
of tumor-stromal interactions in mediating pro-
gression. Two-plus years of experience with mice,
molecular biology, genomics, and cell culture are pre-
ferred. Please send curriculum vitae and two letters of
reference to: **Charles A. Powell, M.D., Department
of Medicine, 630 West 168th Street (Box 91), New
York, NY 10032;** or e-mail: cap6@columbia.edu.
*Columbia University takes Affirmative Action toward Equal
Employment Opportunity.*

Excellence always attracts the best

PROFESSORSHIPS AND EARLY CAREER LECTURESHIPS IN SCIENCE, ENGINEERING AND MEDICINE

University College Dublin (UCD), Ireland's largest and most diverse university, has played a unique role in the development of modern Ireland. Building on this track record, UCD has recently launched a new strategic plan (available on www.ucd.ie) with the explicit goal of establishing UCD as a European leader in 4th level training (masters and PhD training) and research.

To drive this transformation, UCD is inviting expressions of interest from both senior professors and early-career lecturers in the range of disciplines set out below within the context of the recently announced call for proposals from Science Foundation Ireland (the "Stokes Professorship and Lectureship Programme" – available at www.sfi.ie).

BIOLOGICAL SCIENCES

- GENETICS
- BIOCHEMISTRY
- CELL BIOLOGY
- COMPARATIVE GENOMICS
- DEVELOPMENTAL BIOLOGY
- ENVIRONMENTAL SCIENCE
- PROTEOMICS AND PROTEIN SCIENCE
- INORGANIC CHEMISTRY
- SYSTEMS BIOLOGY

BIOMEDICAL AND BIOPHARMACEUTICAL SCIENCES

- MICROBIOLOGY
- PHYSIOLOGY
- PHARMACOLOGY
- PHARMACEUTICAL SCIENCES
- MOLECULAR MEDICINE

- MOLECULAR NUTRITION
- NANOBIOLOGY AND NANOMEDICINE
- NEURODEGENERATION
- DIABETES AND CARDIOVASCULAR RESEARCH

ENGINEERING, MATHEMATICAL AND PHYSICAL SCIENCES

- ADVANCED IMAGING RESEARCH
- APPLIED MATHEMATICS
- BIOENGINEERING
- WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
- COMPUTER ENGINEERING AND EMBEDDED SYSTEMS
- ENERGY RESEARCH
- COMPUTATIONAL SCIENCE AND INFORMATICS
- MATERIALS SCIENCE

Expression of interest enquiries can be emailed to the following address: stokes@ucd.ie from where your query will be routed to the appropriate person.

Closing date for expressions of interest is 4th May, 2007.

Please see our website for further information on these positions

www.ucd.ie/vacancies

UCD is an equal opportunities employer.

IRELAND'S EDUCATION CAPITAL

University College Dublin
An Coláiste Ollscoile,
Baile Átha Cliath

UCD PERSONNEL, UNIVERSITY COLLEGE DUBLIN, BELFIELD, DUBLIN 4, IRELAND.
T:+353-1-716 4900 ; F:+353-1-716 4949

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

Executive Officer
Institute of Medicine

The Institute of Medicine of the National Academies (IOM) seeks an Executive Officer. This is an exceptional opportunity for a strong leader and manager who is well versed in health and government and seeks to make a positive impact on the public's health.

The mission of the Institute of Medicine is to serve as adviser to the nation to improve health. Working outside the framework of government, the IOM provides scientific, evidence-based guidance to policy makers, professionals, civic leaders, and the public at large.

The successful candidate will work closely with the IOM President and will oversee the intellectual, managerial, and financial aspects of the IOM programs in accordance with the policy directives of the IOM President.

This position requires a PhD in a related field or equivalent knowledge with 10 years of related professional experience, 5 of which were in a supervisory capacity.

For more information about this challenging career opportunity and to apply, please visit our website at **http://national-academies.org** - click on Employment - view by department – Institute of Medicine.

The National Academies are nonprofit organizations that work under an 1863 congressional charter, providing independent advice to the nation on matters of science, technology, and medicine. Headquartered in Washington, DC. EOE/M/F/D/V.

national-academies.org

Nebraska Center for VIROLOGY

Faculty Positions in Immunology with an Emphasis on Viral Systems

We are seeking candidates to fill two new faculty positions in immunology with an emphasis on studying viral systems. Appointments may be at the level of Assistant, Associate, or Full Professor and will be in the Nebraska Center for Virology housed in a new state-of-the-art research facility at the University of Nebraska - Lincoln. Excellent opportunities exist for research collaboration between Center faculty located in multiple departments and at other institutions. All candidates should hold a PhD or equivalent. Senior candidates must have a strong research record and current extramural research funding demonstrating expertise in studying the immune response against viral infection. Junior candidates with expertise in broader areas of immunology, who plan to develop a program in a viral system, are strongly encouraged to apply. All candidates will be expected to maintain extramural funding and be involved in graduate and undergraduate teaching. Information about the Center can be found at: **http://www.unl.edu/virologycenter**.

For more information or to apply for this position, visit **http://employment.unl.edu**. To be considered you MUST complete the Faculty/Administrative Information form for Requisition Number 070230 and attach required documents. Review begins **May 15, 2007.**

The University of Nebraska is committed to a pluralistic campus community through affirmative action and equal opportunity and is responsive to the needs of dual career couples. We assure reasonable accommodation under the Americans with Disabilities Act; contact J. Walker at (402) 472-4560 for assistance.

Nebraska UNIVERSITY OF
Lincoln

Sunnybrook
RESEARCH INSTITUTE

Senior Research Associate

Applications are invited for the position of Senior Research Associate. This is a full time position with salary and full benefits commensurate with experience.

The laboratory focuses on the genetics and biochemistry of lymphocyte antigen receptor mediated cellular activation – with a specific focus on the roles of src family tyrosine kinases and the cognate phosphatases that regulate their function. The successful candidate will have demonstrated experience in gene cloning, gene mutagenesis, the creation and utilization of expression vectors, cell lines, T cell clones, primary T lymphocytes and the ectopic expression of kinase and phosphatase variants in these cells.

Curriculum vitae and details of previous research activities along with the names and contact information of three references can be submitted electronically or by mail to:

Michael Julius
Vice President, Research
Sunnybrook Health Sciences Centre
2075 Bayview Avenue, Room A3 33
Toronto, Ontario, M4N 3M5 Canada
michael.julius@sri.utoronto.ca

Sunnybrook Health Sciences Centre and the University of Toronto are committed to employment equity and welcome applications from all qualified women and men, including visible minorities, aboriginal people, persons with disabilities, and persons of a diversity of sexual orientation.

STONY BROOK UNIVERSITY MEDICAL CENTER

Physician/Scientist Positions

Stony Brook University Medical Center is seeking two Physician Scientists for faculty positions at the rank of Assistant Professor, Associate Professor, and Professor in the Department of Medicine in the general areas of infectious diseases and immunity to infection. The candidates will hold a joint appointment in the Department of Molecular Genetics and Microbiology. The Department of Medicine, the Department of Molecular Genetics and Microbiology, and the adjacent Center for Infectious Diseases provide a highly interactive scientific community with a world-class research infrastructure. The School of Medicine and Stony Brook University Medical Center maintain outstanding core research facilities and superb clinical facilities. The position guarantees protected time for research activities. An outstanding start-up package includes laboratory space with start-up funds, excellent benefits, and competitive salaries commensurate with academic qualifications. The candidates are expected to establish or direct an independent extramural research program. This position will require some participation in clinical and teaching activities.

Required: M.D. or M.D./Ph.D., board certification in Infectious Diseases, and be eligible for NYS medical licensure. Priority will be given to candidates with research interests and expertise in the areas of infectious diseases or immune responses to infection to complement or add to existing areas of research strength in both departments.

The review of applications will begin April 2007, and will continue until July 2007, or until the position is filled.

Candidates should send a CV, a brief summary of accomplishments and future research interests (four pages total), and the names and contact information of three references to: Jorge L. Benach, Ph.D., Chair of Search Committee, Department of Molecular Genetics and Microbiology, 130 Life Sciences Building, Stony Brook University, SUNY, Stony Brook, NY 11794-5222

Equal Opportunity/Affirmative Action Employer. Visit *www.stonybrook.edu/cjo* to apply online and for employment information.

HOWARD HUGHES MEDICAL INSTITUTE

HHMI Investigator Competition

The Howard Hughes Medical Institute invites applications for investigator positions from scientists who have demonstrated originality and productivity in biomedical research and who show exceptional promise for future contributions.

Eligibility

:: Ph.D. or M.D. (or the equivalent)

:: Tenured or tenure-track position as an assistant professor or higher academic rank (or the equivalent) at an eligible U.S. institution

:: Four to 10 years of experience since appointment as an assistant professor

:: Principal investigator on one or more active national peer-reviewed research grants with a duration of at least three years

Application deadline: June 13, 2007

Application information:
www.hhmi.org/investigator2008/sci

HHMI investigators are among the most creative and promising biomedical scientists in the nation. They rigorously pursue significant biological questions, develop innovative research tools and methods, and lead their scientific fields into new areas of inquiry. Working to push the boundaries of fundamental knowledge and ultimately to improve human health, HHMI investigators forge links between basic biology and medicine.

The Institute seeks to appoint to its investigator program approximately 50 outstanding scientists. This competition will enable HHMI to strengthen its community of researchers and bring innovative approaches to the study of biological problems not only in the biomedical disciplines but also in adjacent fields, such as chemistry, physics and biophysics, biomedical engineering, and computational biology. Candidates should apply directly to HHMI; prior institutional endorsement is not part of the application process.

HHMI, a nonprofit medical research organization, plays a powerful role in advancing biomedical research and education in the United States. The investigator program rests on the conviction that scientists of exceptional talent, commitment, and imagination will make fundamental biological discoveries for the betterment of human health if they receive the resources, time, and freedom to pursue challenging questions. The Institute's investigators, selected through rigorous national competitions, include 11 Nobel Prize winners and 115 members of the National Academy of Sciences.

The Howard Hughes Medical Institute is an equal opportunity employer. Women and members of racial and ethnic groups traditionally underrepresented in the biomedical sciences are encouraged to apply.

HHMI
HOWARD HUGHES MEDICAL INSTITUTE

Kingdom of Saudi Arabia
Ministry of Higher Education
Taif University

College of Medicine and Medical Sciences and College of Pharmacy

Academic Staff Vacancies for the Academic Year 2007-2008

College of Medicine and Medical Sciences, Taif University, Saudi Arabia, is a newly established institution of higher education that is striving for worldwide recognition, excellence of services and dedication to the advancement of medical science and research. In addition, a pharmacy school, Pharm D program, has started in 2006 under the administration of the medical school.

The college is looking for highly motivated outstanding individuals to fill the vacant academic positions of Professor, Associate Professor or Assistant Professor or Lecturer in the following disciplines to undertake teaching responsibilities as well as research endeavours in the colleges of medicine and medical sciences and pharmacy:

College of Medicine and Medical Sciences:
- **Professor Posts:**
  One position in each for the following specialities: Medicine, Surgery, Family Medicine, Community Health Sciences and Pediatrics.
- **Associate Professor Posts:**
  One position in each for the following specialities: Epidemiology, Obstetrics and Gynecology, Communication Skills, Physical Therapy. Two positions for Histopathologists.
- **Assistant Professor Posts:**
  One position in each for the following specialities: Clinical Psychology, Community Health Sciences, Family Medicine, General Surgery, Haematology (clinical), Internal Medicine, Laboratory Medicine, Medicine (Endocrinology), Psychiatry, Clinical Radiology, Otolaryngology, General Pediatrics, Neurosurgery, Medicine (Neurology), Urology (Medicine), Pediatric Pulmonology, Pediatric Gastroenterology, Orthopedics, Anesthesiology, Obstetrics and Gynecology, Cardiology, Oncology, Nuclear Medicine, Physical Therapy, Biostatistics and Epidemiology.
  Two positions in each for the following specialities: Intensive Care Medicine and Environmental Health.
  Three positions for English for Medical Sciences.
- **Lecturer Posts:**
  Five positions are open for lecturers in Basic Medical Sciences and one position in Medical Ethics.

College of Pharmacy:
- **Professor, Associate Professor or Assistant Professor Posts:**
  One position in each for the following specialities: Clinical pharmacy, pharmaceutical Microbiology and Physiology.
  Two positions in Pharmaceutics and Pharmaceutical Technology.
- **Lecturer Positions:**
  Two positions for English Language Lecturers.

Qualifications: PhD or equivalent for Professor, Associate Professor or Assistant Professor positions, unless otherwise specified. MSc for English Language Lecturers and MBBS for the Lecturer positions at the College of Medicine.

Experience: Teaching experience in a recognised university (At least ten years for Professor, five years for Associate Professor, two years for Assistant Professor or Lecturer). Scientific productivity must be measured in form of publications in reputable peer reviewed scientific journals.

The university benefits package includes competitive salaries, housing, medical coverage, 60 days annual leave, about 20 days holidays, airline tickets, relocation allowance and end of service gratuity. Academic rank and salary will be commensurate with qualifications.

Interested individuals are requested to submit their applications, which should include an updated detailed C.V., list of publications and names and addresses of three referees. Complete applications should be addressed to:

The Dean
College of Medicine and Medical Sciences
Taif University
P O Box 888
Taif 21944
Kingdom of Saudi Arabia

Email: **taam_tm@yahoo.com, doc_tm100@hotmail.com**
Fax: +966-2-7250528

---

ulm university   universität
u ulm

The **Medical Faculty of the University of Ulm** invites applications for a position of a

Professor (W 3)
of Molecular Aging Research

at the **Department of Dermatology and Allergology**
– Head of Department: Professor Dr. Karin Scharffetter-Kochanek

Applications are invited from internationally recognized scientists for the position of a full Professor of Aging Research within the Medical Faculty of the University of Ulm.

Applicants are expected to have a documented record of accomplishment in the field of aging and age-related diseases. Applicants from other research fields are also encouraged to apply but must have a strong interest to extend their research track into the field of aging research. The successful candidate is an outstanding scientist, with experience in areas that could include cell/molecular biology, biochemistry, genomics and genetic model systems to study the molecular basis of cellular and organismal longevity and senescence.

The position is embedded in the clinical research group "Molecular and cellular aging – from mechanisms to clinical perspective" funded by the German Research Foundation (DFG). The German Research Foundation expects the candidate to apply for funding of a project that is within the aims of the clinical research group, and to develop strong independently funded research programs to complement and enhance the already existing strengths in aging research. The candidate with a successful proposal will be equipped with a competitive start up package.

Requirements include a PhD or MD degree, postdoctoral experience, and a postgraduate lecture qualification or an equivalent to German "Habilitation". Applicants should have a sound research record and a successful track record in grant acquisition.

The position is tenure-track. With the first appointment, the contract of employment will be limited for the funding period of the clinical research group. Subsequently, tenured employment is possible, and will depend on an evaluation.

The University aims at increasing the representation of women among the faculty staff and therefore explicitly encourages applications from female scientists.

Applications including the application form, a full CV, a complete list of grants and publications, and a summary of current and proposed research activities and objectives should be sent and received not later than May 15, 2007 to: University of Ulm, Professor Dr. K.-M. Debatin, Dean of the Faculty of Medicine, D-89069 Ulm, Germany.

The application form is available at http://www.uni-ulm.de/medizin/ (Fakultät, downloads, Fakultät, Bewerbungsformular.doc)

The University of Ulm is an equal opportunity employer and welcomes applications from all qualified individuals.

Does your next career step need direction?

I have a great new research idea. Where can I find more grant options?

For a career in science, I turn to Science

ScienceCa

We know science

You know, ScienceCareers.org is part of the non-profit AAAS

That means they're putting something back into science

Looking for solid ground in the ever-changing landscape of science & technology policy and budget issues?

Join the nation's top S&T experts at the 32nd Annual AAAS Forum on Science & Technology Policy
3–4 May 2007 • Washington DC
International Trade Center in the
Ronald Reagan Building

The AAAS Forum on Science and Technology Policy provides a setting for discussion and debate about the federal budget and other policy issues facing the science, engineering, and higher education communities. Initiated in 1976 as the AAAS R&D Colloquium with about 100 participants, the Forum has emerged as the major public meeting in the U.S. devoted to science and technology policy issues. It annually draws upwards of 500 of the nation's top S&T policy experts.

- Get a full analysis of the President's federal R&D funding proposals.

- Have an opportunity to meet directly with key S&T policymakers.

- Learn how the changes in Congress are affecting S&T policy issues.

- Network with colleagues, including top decisionmakers in science and technology policy from all sectors.

- Learn about broader national and international developments that will affect strategic planning in universities, industries, and government.

- Registrants will receive, at the Forum, *AAAS Report XXXII: Research and Development, FY 2008*, a comprehensive analysis of the proposals for the FY 2008 budget, prepared by AAAS and a group of its affiliated scientific, engineering, and higher education associations.

For more complete details on the program, hotel registration and on-line registration, please visit the website: www.aaas.org/forum.

www.aaas.org/forum

AAAS
ADVANCING SCIENCE. SERVING SOCIETY

ScienceCareers.org